

Shears, Beverly

73498

From: Devi, Sarvamangala
Sent: Friday, August 16, 2002 11:16 AM
To: Shears, Beverly
Subject: 09/921,157

Good morning Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 3 in application SN 09/921,157?

Thanks.

S. DEVI, Ph.D.
AU 1645
CM1-7E15

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:27:24 ; Search time 42.28 Seconds
(without alignments)
3404.722 Million cell updates/sec

Title: US-09-921-157-3

Perfect score: 6671

Sequence: 1 MEIQQTHRKINRPLVSLV.....HNLISNIGHFASNLGMYRSF 1296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	6554	98.2	1296	14	AA41198	CT, Helicobacter
2	6300.5	94.4	1287	16	AA79944	Helicobacter pylori
3	6220	93.2	1290	19	AA98269	H. pylori GHPO 374
4	6108	91.6	1288	18	AAW5547	H. pylori ORF 14ee
5	6108	91.6	1288	18	AAW55685	H. pylori ORF 07ee
6	2707.5	40.6	578	21	AAW52575	Helicobacter pylori
7	1792.5	26.9	407	18	AAW5290	H. pylori ORF 02ep
8	1792.5	26.9	407	18	AAW20652	H. pylori derived
9	1694.5	25.4	513	16	AA79945	Helicobacter pylori
10	1538.5	23.1	326	21	AAW52500	Helicobacter pylori
11	1406	21.1	288	18	AAW27720	H. pylori Vaca pro

12	1198	18.0	257	18	AAW55462	H. pylori ORF 07ap
13	1052	15.8	216	20	AAW89963	Antigen from clust
14	1029.5	15.4	224	21	AAW52506	Helicobacter pylori
15	861.5	12.9	176	18	AAW55346	H. pylori ORF 07ep
16	824	12.4	211	18	AAW55324	H. pylori ORF 07ap
17	824	12.4	211	18	AAW20755	H. pylori protein.
18	674	10.1	1484	22	AAW55686	H. pylori ORF 07ee
19	674	10.1	2902	22	AAW46351	H. pylori HPN165 p
20	648.5	9.7	2893	19	AAW98828	H. pylori GHPO 148
21	648.5	9.7	2893	19	AAW71556	Helicobacter poly
22	626	9.4	1382	19	AAW11001	H. pylori ORF 1lee
23	623	9.3	1413	18	AAW20725	H. pylori secreted
24	581	8.7	1974	19	AAW98391	H. pylori GHPO 57
25	569.5	8.5	786	18	AAW55464	H. pylori ORF 07ap
26	536	8.0	133	18	AAW55437	H. pylori ORF hp5p
27	509	7.6	560	18	AAW20409	H. pylori secreted
28	504	7.6	141	18	AAW20260	H. pylori protein.
29	504	7.6	141	18	AAW24621	H. pylori protein.
30	486	7.3	458	18	AAW55372	H. pylori ORF hp2e
31	486	7.3	458	18	AAW20964	H. pylori cytoplas
32	477.5	7.2	1213	18	AAW55735	H. pylori ORF 07ee
33	471	7.1	92	21	AAW52604	Helicobacter pylori
34	431	6.5	95	18	AAW55322	H. pylori ORF 07ap
35	431	6.5	95	18	AAW20754	H. pylori protein.
36	427	6.4	90	18	AAW20321	H. pylori protein.
37	421	6.3	93	18	AAW20362	H. pylori protein.
38	386	5.8	264	18	AAW20199	H. pylori cytoplas
39	320.5	4.8	1612	19	AAW65088	R. prowazekii S-la
40	289	4.3	1221	21	AAW01825	Haemophilus Influe
41	289	4.3	1227	21	AAW01824	Haemophilus Influe
42	288.5	4.3	1601	18	AAW30292	Non-typeable Haemo
43	287.5	4.3	1338	14	AAW41731	High molecular wei
44	287.5	4.3	1529	14	AAW41732	High molecular wei
45	284.5	4.3	1598	18	AAW30291	Non-typeable Haemo

ALIGNMENTS

RESULT 1
AA41198
ID AAR41198 standard; Protein; 1296 AA.

XX AAR41198;

XX DT 17-MAR-1994 (first entry)

XX DE CT.

XX Cytotoxin; CT; H. pylori; precursor; vacuolation; cell death; hsp60;
KW heat shock protein; cytotoxin-associated immunodominant antigen; CAI;
KW hsp; type B; gastritis; peptic ulcer; eukaryote; gastric tumours.

OS Helicobacter pylori.

XX Key Location/Qualifiers

FT Peptide 1..33

FT Protein /note= "Signal peptide"

FT Protein 34..1296

XX /note= "Mature Ct"

PN W09318150-A.

XX 16-SEP-1993.

XX PF 02-MAR-1993; 93WO-EP00472.

XX PR 02-MAR-1992; 92IT-OFI0052.

XX PR 25-JAN-1993; 93WO-EP00158.

XX PA (BIOC-) BIOGINE SCLAVO SPA.

XX Bugnoli M, Covacci A, Macchia G, Rappuoli R, Telford J;

XX WPI; 1993-303464/38.
 DR N-PSDB; AAQ48732.
 XX
 PT Recombinant Helicobacter pylori protein and corresp. gene - is a
 PT cytotoxin, antigen or heat shock protein used for treating and
 PT preventing type B gastritis, gastric ulcers and gastric tumours
 XX
 PS
 XX Claim 3; Fig 2; 83pp; English.
 CC This sequence represents the precursor form of cytotoxin, CT, of H.
 CC pylori. The DNA encoding this protein was isolated from two
 CC overlapping fragments corresponding to about 10 kb of the H. pylori
 CC genome. This protein has a molecular weight of 139.8 kD and serves as
 CC a precursor to a protein having an approximate weight of 100 kD and a
 CC cytotoxic activity. The amino acid positions 34-56 indicate a
 CC previously isolated N-terminal peptide. This sequence is preceded by
 CC 33 amino acids which resemble prokaryotic leader sequences, thus this
 CC sequence is likely to represent the N-terminal of the mature protein.
 CC A previously isolated cytotoxin of 87 kD may be produced by further
 CC processing or proteolytic degradation during purification. The mature
 CC cytotoxin causes vacuolation and death of a number of eukaryotic cell
 CC types. This protein, and others derived from H. pylori, esp.
 CC cytotoxin-associated immunodominant (CAI) antigen or a heat shock
 CC protein (hsp) (see also AAR41199-200), may be used to treat, prevent and
 CC diagnose H. pylori infection. H. pylori is the causative agent of type
 CC B gastritis, peptic ulcers and gastric tumours.
 XX
 XX Sequence 1296 AA;

Query Match 98.2%; Score 6554; DB 14; Length 1296;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1277; Conservative 2; Mismatches 17; Indels 0; Gaps 0;

Qy 1 MEIQOHRKINRPLVSLALVGLVSLVITPOQSHAAFFTVIIPALVGGTACTVAGTVSGL 60
 Db
 Qy 1 melqqrhrklnrplvslalvglvslvtpqshaaftvtiipalvvggtactvagtvsyl 60
 Db
 Qy 61 LSWGLKQAEANKTPDKPWRIQAGKGFNEFPNKEYDYRLSLSSKIDGWDGWNAAAR 120
 Db
 Qy 61 lswglkqaeanktpdkpdkwriqagkgfnefpnkeydyrlslsskldgwdgwnaar 120
 Qy 121 HYWYKGGQONKLEVDMDKAVGTYTLGLRNFTGGDLVNMOKATLRLQFNGNSFTSYKD 180
 Db
 Qy 121 hywykggqonklevdmkdavgttytlglrnfthgdlvnmokatlrlqfngnsftsykd 180
 Db
 Qy 181 SADRTTRVDFNAKNISIDNFEINRNVCAGKASSTVLTQASEGTTSDKNAEISLYD 240
 Db
 Qy 181 sadrttrvstlksqililqkstrvgsggrkasstvtlqasegitsdknaeisllyd 240
 Qy 241 GATNLASSSVKLMGNVMMGLQVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
 Db
 Qy 241 gatnlasssvklmgvnmgrlqvaylapyystintskvtgevnfnhlvtvgdknaaqa 300
 Qy 301 GIANKTNTICTLDWQAGLNIITAPPEGYKDKPNTPSOSGAKNDKNSAKNDKQESS 360
 Db
 Qy 301 gilanckntnigtldwqaglniitappeggykdkpnntpsosgakndknesakndkqess 360
 Qy 361 QNNSTOVINPPNSAQKTEVQPTQIDGPFAGGKDTVVNIRNTNADGTLRVGGFKASL 420
 Db
 Qy 361 qnnstovinnppnsaqktevqptqidgpfaggkdtvvnnrntnadgtlrvvggfkasl 420
 Qy 421 TTNAHLHIGKGVNLSNQASGRSLIVENLTGNTVDGPLRVNNQVGGYALAGSANFEF 480
 Db
 Qy 421 ttnaahlhigkgvnlslngqasgrslivenltgntvdtgplrvnnqvgyalagsanfef 480
 Qy 481 KAGTDTNCTATFNNDISLRFPVNLKVDATNFAGKIDTGNCGFNTLDFSCVTDKVINK 540
 Db
 Qy 481 kagtdtngtatfndislsgrfpvnlkvdantfnagkidtgnngfntldfsgvtdkvink 540
 Qy 541 LITASTNVAVKNFINELIVKTNGISVGEYTHFSEDTGSSQSRINTVRLTGTSLFSGGV 600
 Db

Db 541 litastnvavknfninelivktngisvgeythfseidlgsgsrintvrltgrslfsggv 600
 Qy 601 KFKGGEKLIVIDEFYSPWNYFDARNIKNVEITNKLAFCQSPMGTSKLMFNNTLQGNA 660
 Db
 Qy 601 kfkgeklividefyyspwnyfdarniknveitnklafcqspmgtsklmfnntlqgna 660
 Qy 661 VMDYSQSNTUIGDFTINNOGTFINLVGRGKVATLSVGNAAAMFNNDIDISATFYKPLI 720
 Db
 Qy 661 vmdysqfntuigtfdtinnogtfinnlvgrgkvatlsvgnaaamfnndidsatfykpli 720
 Qy 721 KINSAQDLINKTEHVLKAKTIIGYGNVSTGTNGISNVNLEQPKERLALYNNRMMDTCV 780
 Db
 Qy 721 kinsaqdlinktehlkaktiigygnvstgtngisnvnleeqkerlalynnrmmdtcv 780
 Qy 781 VRNTDDIKACGMAIGDQSMVNNPDNYKILIGKAWKNIGISKTANGSKISVYVYLGNSPTPE 840
 Db
 Qy 781 vrntddikacgmaigdqsmvnnpdnykyligkawknigisktangskisvvyvlgnspte 840
 Qy 841 NGGNTTNPNTTNSARSNANALQAQAPQPSATPNLVAINQHDFTGIESVFELANRSK 900
 Db
 Qy 841 nggnttnpntttnsnarsannalaqnapfaqpssatpnlvainqhdftgiesvfelanrsk 900
 Qy 901 DIDTLYANSQAQGRDLLQTLIDSHDAGYARKMIDATSNSEITKOLNTATTTLNIAASLE 960
 Db
 Qy 901 didtlyansgaqgrdllqtlidshdagyarkmidatsaneitkolntatttlniasle 960
 Qy 961 HKTSGLQTLSLSNAMILNSRLVNLRRHTNHIDSFARLQALQKDOKFASLESAAEVLQF 1020
 Db
 Qy 961 hktsglqtlslsnamilnsrlvnlrrhtnhidsfakrlqalkdokfasesaaevlyqf 1020
 Qy 1021 APKYEKPTNWANAIGTSTLNGSGNASLYGTSAGVDAYLNGQVEAIVGGFGSGYSSEFNN 1080
 Db
 Qy 1021 apkyekptnwanaigtstlngsgnaslygtsagvdaylngqveaivggfsgyssefnn 1080
 Qy 1081 RANSLNSGANTNFVGSYRIFANQHEPFAQAGALGSDQSLNFKSALLODLNOSYHYLA 1140
 Db
 Qy 1081 ranslmsgantnfvgysrifanqhhepfaogalgdsqslnfksalldlnosyhyyla 1140
 Qy 1141 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
 Db
 Qy 1141 ysaatrasgydfaffrnaivlpkpsgvsynhlgstnfksnstnovalkngsssqhlfna 1200
 Qy 1201 SANVEARYYGDTSYFYMNAGVLOEFAHVGVSNNAAASLNTFKVNAARNPLNTHARVMGGE 1260
 Db
 Qy 1201 sanvearyygdtsyfymnagvlqefahvgvsnnaaaslntharvmmgge 1260
 Qy 1261 LKLAKEVFLNLGVVYLNHLNLSNIGHFASNLGCMRYSF 1296
 Db
 Qy 1261 lklakevflnlgvvylhnlslnighfasnlgmrysf 1296

RESULT 2
 AAR79944
 ID AAR79944 standard; Protein; 1287 AA.
 XX
 AC AAR79944;
 XX
 DT 26-MAR-1996 (first entry)
 XX
 DE Helicobacter pylori vacuolating toxin.
 XX
 KW Vacuolating toxin; vaccine; immunisation; therapy; mutant;
 KW Infection; Helicobacter pylori.
 OS
 OS Helicobacter pylori.
 PN
 PN WO9522988-A1.
 XX
 PD 31-AUG-1995.
 XX
 PF 23-FEB-1995; 95WO-US02219.
 XX
 PR 23-FEB-1994; 94US-0200232.

DR N-PSDB; AAX13988.

XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases

XX

PS Claim 8; Page 273-278; 2054pp; English.

XX

CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.

XX

SQ Sequence 1290 AA;

Query Match 93.2%; Score 6220; DB 19; Length 1290;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1207; Conservative 38; Mismatches 45; Indels 6; Gaps 2;

QY 1 MEIQOTHEKINRPLVSLALVGLYSITPQOSHAAFFTTVIIIPAIVGGIATGATVGTSGSL 60
DB 1 melqgthrklnrpvlslalvglvsltpqgshaaffttviiipaivvggiatgavgtvsgl 60

QY 61 LSWGLQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGWAAR 120
DB 61 lgwglqaeaeanktpdkpkvwrliqagkgfnefpnkeydlyrslsskidgwdgwnaat 120

QY 121 HYWKGGQONKLEVDMDKAVGTYTILSLGRNFTGGDLVNMOKALRLGQFNGNFTSYKD 180
DB 121 hywkggqonklevdmdkavgtvtyttilslgrnftggdlvnmokalrlgqfngnftszykd 180

QY 181 SADRTTRVDFNAKNISIDNFEIINRVGSGAGRKASSVLTQLQASEGITSKNAEISLYD 240
DB 181 sadrttrvdfnknislidnfeiinrvvgsgagrkassvltqlqasegitssknaeislyd 240

QY 241 GATNLASSSVKLMGNVMGRLOXVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
DB 241 gatnlaasssvklmgvnmgrloxvgaylapstintskvtgevnfnhlvtvgdknaqa 300

QY 301 GIIANKRTNIGTLDLWQAGLNIITAPPEGGYKDKPNTPPSQSGAKNDKNEAKNDKBESS 360
DB 301 giiankrtnigtldlwqaglniitappeggykdkpnpnpssqsgakndkneakndkbess 360

QY 361 QNNSNTQVINPNSAQKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTTIRVGGFKASL 420
DB 361 qnnsntqvinnpnsaqktevqptvidgpfaggkdtvvnninrntnadgttirvggfkasl 420

QY 421 TTNAHLHIGKGVNLSNQASGRSLIVENLTGNTVDGPLRVNMQVGYALAGSSANFEF 480
DB 421 ttnahlhigkgvnlslsqasgrslivenltgntvdgplrvnmqvgyalagssanfef 480

QY 481 KAGDTTKNGTATFNNDISLGRFVNLKVDYAHANFKGIDTGNCGNTLDFSGVTDKVNLNK 540
DB 481 kagdttkngtatfnndislgrfvnlkvdyahankgidtgnngntldfsgvtdkvnlknk 540

QY 541 LITASTNVAVKFNFINELIVKTINGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV 600
DB 541 litastnvavkfnfinelivktingisvgeythfseidigsqsrintvrltgtslfsggv 600

QY 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLFAGPQSGPWGTSKLMFNNTLQONA 660
DB 601 kfkgeklvldfyyspwnyfdarniknveitnklfagpqsqpwgtsklmfnntlqona 660

QY 661 VMDYSQFSLNLTQDGFINNQCTINLYRGKAVTILSVGNAAAMFNNDIDSATGFYKPLI 720
DB 661 vmdysqfslnltqdgfinnqctinlyrgkavttilsvgnaaamfnndidsatgfykpli 720

QY 721 KINSAQDLINKTEHYLLKAKIIGYCNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 780
DB 721 kinsaodlikntehyllkakiiigycnvstgtngisnvnleqfkerlalynnnrmdtcv 780

DB 716 kinsaodliknkehvllkakliigynvslgcnslsnvnlleqfkerlalylnnnrmdlcv 775

QY 781 VRNTDDIKACGMAICDQSMVNNPNYKYLIKAWKNIGISKITANGSKISVYVLGNSTPTE 840
DB 776 vrntddikacgtaignqsmvnnpnykylikawkngigistangskislvyvlgnstpte 835

QY 841 NGGNTNLPTNTTNSARSNANALAQNAFPAPSPATPNLVAINQHDFTGFTIESVFELANRSK 900
DB 836 kggntnlptnttnsarsannalagnapfaqsatpnlvainqhdftgiesvfelanrsk 895

QY 901 DIDTLYANSAGQGRDLQTLTLLIDSHDAGYARKMIDATYANEITKQLNTATTTLNNIASLE 960
DB 896 didtlyansagqgrdlqltllidshdagyarqmdntstgetkqlnaatatlanniasele 955

QY 961 HKTSGLOTLSLSNAMIINSLRNLNLSRRHTNHIDSFARKLQALQDKQKFALESAAEVLQYF 1020
DB 956 hktsslqtlslsnamiliinlslrrhtnnidafaqlqalkqkfaslesaeavlyqf 1015

QY 1021 APKYEKPTNWWANAIGGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSGYSSPNN 1080
DB 1016 apkyekptnwwanaiggtslngngnaslygttsagvdaylngveavlgvgfsgygyssfn 1075

QY 1081 RANSLNCANNNTNFGVSRIFANOHEPDEFAQAGALGSDOSLSLNFKSLALQDLNOSYHYLA 1140
DB 1076 qanslncannntnfgvsrifanohedfdeagagalsdqsslnlfnksalrlngdsynyla 1135

QY 1141 YSAATRASGYDFAFFRNALVLPKPSVGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1200
DB 1136 ysaatrasgydfaffrnalvlpkpsvgvsnhlgstnfksns-nqvalkngsssqhlfna 1194

QY 1201 SANVEARYYIGDTSYFYMNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
DB 1195 sanvearyyyigdtstyfyminagvloeafaahvgssnnaaslntfkvnaarhplstharvmgge 1254

QY 1261 LKLAKEVFLNLGVVYLHNLISNIGHFASNLGMRYSF 1296
DB 1255 lklakevflnlgvylhnlisnighfnisnighmrysf 1290

RESULT 4
AAW5547
ID AAW55547 standard; Protein; 1288 AA.
AC AAW55547;
XX
DT 24-JUN-1998 (first entry)
XX
DE H. pylori ORF 14ee41924_2458267_c2_93 secreted protein.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterie; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
XX
PN W09737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97WO-US05223.
XX
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
PA (ASTR) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI; 1997-503122/46.
DR N-PSDB; AAV24956.

XX Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Disclosure; Page 750-753; 1145pp; English.
XX
CC This sequence is a H. pylori secreted protein. The protein may be used
CC in a vaccine to prevent or treat H. pylori infection or to identify
CC H. pylori polypeptide binding compounds, useful as potential H. pylori
CC life cycle activators or inhibitors. The DNA and probes derived from it
CC may be used for the identification of H. pylori in a sample and the
CC diagnosis of H. pylori infection. Nucleic acid sequences complementary
CC to the DNA act as antisense sequences and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can be
CC used in immunoassays to evaluate the abundance and distribution of
CC H. pylori-specific antigens. The genomic sequence of H. pylori
CC (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 1288 AA;

Query Match 91.6%; Score 6108; DB 18; Length 1288;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1181; Conservative 53; Mismatches 54; Indels 8; Gaps 2;

Qy 1 MEIQTHRKINRPIVSLALVGNALVSIPTQOQSHAAFTTVPPIAVGGTATGATVTSGL 60
Db 1 meiqthrkirplvslalagalisabqeshaaftftvpiavvgiatgtavgtvsgl 60
Qy 61 LSWGLQAEANKTPDKPKWRIQAGKGFNPFKEVDLYRSLSSKIDGMDGNNAAR 120
Db 61 lswglkqaeanktpdkpkwriqagkgnfepkneydlyrslsskidgmdgwnaar 120
Qy 121 HYWKGQONKLEVDMDKAVGTYTSLGRNFTGGDLVNMQKATLRLQGFNGNSFTSYKD 180
Db 121 hywkgqwnklevdmdkavgtyklsgrnftggdlvnmqkatrlrgqfngnsftsykd 180
Qy 181 SADRTTRVDFNAKNISIDNFVNNRVSGAGRKASSVLTLOASEGITSKNAEISLYD 240
Db 181 sadrttrvfnaknisidnfvnnrvsgagrkassvtltloasegitssknaeislyd 240
Qy 241 GATLNLASSVKLMGNVWMLQYGVAYLAPSYSTINTSYTEVNFHNLTVGDKNAQA 300
Db 241 gatlnlansvklngnvmwmlrlyqvgaylapsystintskvgevdhltvtvgdqaqa 300
Qy 301 GIANKTNIGTLDLQWAGLNIAPPPEGKYDKPNNTPSQSGAKNDKNEAKNDQESS 360
Db 301 giianktkhtgldlqwsaglniapppegkydkpnsttsqsgtkndkkeisqn----- 354
Qy 361 QNNSNTQVINPPNSAQKTEVQPTQVIDGPFAGGKDTVVNNINRINTNADGTIRVGFKASL 420
Db 361 -nnsntevinnpntqketeptqvidgpfaggkdtvnnihntkadtikvvgfkasl 413
Qy 421 TTNAHLHIGKGVNLSNOAGRSILVNLGTNIVDGPLRVNNOVGGVALAGSSANFEF 480
Db 421 ttnaahlhigkgvnlsnqasgrtllvnlgtndgplrvnnqvvgvalagssanfef 473
Qy 481 KAGDTKNGTATFNNDISILGRFVNLKVDAAHTANFKGIDTNGGFFNTLDFSGVTDKVNINK 540
Db 474 kagvtdkngtatfnndisilgrfvlkvdtahtanfkidtgngffntldfsgvtdkvnink 533
Qy 541 LTASTNVAVKFNINELIVKTNIGTSVGEYTHFSSEIDISQSRINTVRLTGTFRSLFSGV 600
Db 541 ltastnvavknfninelivktnigtsvgeythfsseidissqsrintvrltgrslfsgv 600

Db 534 litastnvavknfninelivktnigtsvgeythfsseidissqsrintvrltgrtsifsgv 593
Qy 601 KFKGGEKLVIDFEFYSPWNYPDARNIKNVEITNKLAFGPQSGPWGTSKLMFNLLTGONA 660
Db 594 kfksgelivndfyyspwnyfdarnvknvelcrkfasstpepwtgtsklmfnlltgona 653
Qy 661 VMDYSQFSLNTIQGDFINNOGTINTYLVRGGKVATLSVGNAAAMFNNDISATGYKPLI 720
Db 654 vmdysqfslntiqgdfinnogtintylvrggkvatlnvgnaaamfnndidsatgyfkpli 713
Qy 721 KINSQADLIKNTHEHLLAKAKIIGYGNVSTGTNGISNVNLEEFKRLALYNNNNMDTCV 780
Db 714 kinsqadlikntehvllkakiliygnvstgtngisnvnlleefkerlalynnnnmdtcv 773
Qy 781 VRNTDDIKACGAIQDSQVMVNPNDYKYLIGKAWKNIGISTANGSKTSVYVYLGNSPTPE 840
Db 774 vrntddikacgaimgnqsmvnpndnykyligkawnigisktangsktsvvyvlgnsptpe 833
Qy 841 NGGNTNLTPTNTSARNALQAQNAFPAQSPATPNLVAIINOHDFGTITESFELANRSK 900
Db 834 nggntnltptntnnaahsanyalvknafah-satpnlvainqhdfgtiesvfelanrsk 892
Qy 901 DIDTLIYANGSAGRDLLQTLTLLIDSHDAGYARKMIDATSAEITKQLTATTTLNIAELE 960
Db 893 didtlythsgagrdllqtlidshdagyardmtdntstgeitkqlnaatdalinvasle 952
Qy 961 HKTSGLTQLSNAMLNSRLNRRHTNHIDSFARKLQALKDOKFASLESAAAEVLYOF 1020
Db 953 hkqsgltqlslsnamlnsrlvnlrrkhtnhinsfaqralkqgefasesaaevlyof 1012
Qy 1021 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGSSFN 1080
Db 1013 apkyekptnvwanalggasnlsgsnaslygttsagvdaflngvnealvggfygssfn 1072
Qy 1081 RANSLNSCANNTNFGYSRIFANQHEFDEAOGALGDSOSSLNFKSALLQDLNQSHYLA 1140
Db 1073 qanslsgannanfgystrffanqhefdeagdgalsdsslnfkstllqdlngshyyla 1132
Qy 1141 YSAATRASYGYDFAFRNALVLKPSGVSYNHLGTSNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1133 ysatarasygydfafrnalvlkpsgvsyhlgstnfksnsgvalkngassqhlfna 1192
Qy 1201 SANVEARYYGGTSTYFYNAGVQLQEFHGVHVSNNAAASLNTFKYNAARNPLNTHARVMGGE 1260
Db 1193 nanvearyyggtsfyhagvqlqefhgvshndvaslntfkinaarsplstiyarammge 1252
Qy 1261 LKLAKEVELNLGVVYVHLNLSINIGHFASNLGMRYSF 1296
Db 1253 lqlakevfinlgvvyvhlhlnslshfasnlgmrysf 1288

RESULT 5
AAW55685
ID AAW55685 standard; Protein; 1288 AA.
XX
AC AAW55685;
XX
DT 07-JUL-1998 (first entry)
XX
DE H. pylori ORF 07ee11402_2458267_c3_108 cell envelope OMP.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
OS Helicobacter pylori.
XX
PN WO9737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97WO-US05223.
XX

XX Helicobacter pylori; two-hybrid system; protein-protein interaction;
KW bait polypeptide; gastric ulcer; antibacterial.
XX Helicobacter pylori.
OS Helicobacter pylori.
XX WO2000066722-AL.
XX 09-NOV-2000.
XX 14-APR-2000; 2000WO-IB00603.
XX 30-APR-1999; 99EP-0401066.
XX (HYBR-) HYBRIGENICS SA.
PA Legrain P, Selig L, Rain J;
PI WPI; 2000-687535/67.
XX N-PSDB; AAC97321.
DR A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
cell clones -
XX Example 5; Page 202-204; 267pp; English.
XX The present sequence is a bait polypeptide used in a Helicobacter
XX pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX Sequence 578 AA;
Query Match 40.6%; Score 2707.5; DB 21; Length 578;
Best Local Similarity 90.1%; Pred. No. 7.8e-149;
Matches 525; Conservative 27; Mismatches 26; Indels 5; Gaps 1;
QY 182 ADRTTRVDFAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYDG 241
DB 1 adrttrvdfnakilidnfeinnrvsgagrkasstvtlqasegitsknaeislydg 60
242 ATLNLAASSVKLMGNVWGRLOYGAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQAG 301
DB 61 atlnlasvsklmgnvwmgrloygaylapsystintskvtgevnfnhnlvtgdhnaaqag 120
QY 302 ITANKTNTGTLDMOSAGLNIAPPEGYKDKPNWTPSQSGAKNDKESAKNDKQESSQ 361
DB 121 iasankthlgtldlwsaglniappegykdkpdkps-----nttqnannngnpsaq 175
QY 362 NNSNTQVINPPNSAQTEVQPTQVIDPAGGKDTVNNRINRINADGTRVGGFKASLT 421
DB 176 nnsntqvinppnsaqteiqptqvidpaggkdtvnnidrtntnadgtikvgykaslt 235
QY 422 TNAHLHICKGVNLSNOASGRSLIVENLTGNTVDGPLRVNNOVGYYALAGSSANFEFK 481
DB 236 tnaahlhikgginlsnqasgrtllivnltgnitvdgplrvnnqvggylagssanfek 295
QY 482 AGTDTKNGTATFNNDISLGRFVNLKYDAHTANFKGIDTNGGFGENTLDFSGVTDKVNINKL 541
DB 296 agtdtkngtatfndndislrgrfvlkvdahntanfkidtgngfntldfsgvtgkvninkl 355
QY 542 ITASTWAVKFNINELIVKTNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGVK 601
DB 356 itastwvknfnlnelivktnngsvgeythfsedigsqsrintvrltgttrslfsggvk 415

QY 602 FKGEKLVIDEFYISPNWYFDARNIKNVEITNKLAFGPOGSPMGTSKLMFNLTIGONAV 661
DB 416 fksgeklvidefyispnwfyfdarniknveitrkfastpenwgtksklmfnltlqgnav 475
QY 662 MDYSQFSNLTIOGDFINNOGTINYLVRGKGVATLSYGNAAAAMFNNDIDSATGYKPLIK 721
DB 476 mdysqfsnltioqdfinnogtlnylvrggvatlnvgnaaamfnsnvdsatgyqplmk 535
QY 722 INSAQDLIKNTBHVLLKAKIIGYGNVSTGTNGISNVNLEEQFK 764
DB 536 insaqdliknkvhllkakliigygnvslgtnsnsnvnlieqfk 578
RESULT 7
AAW55290
ID AAW55290 standard; Protein; 407 AA.
XX AC AAW55290;
XX DT 15-JUN-1998 (first entry)
DE H. pylori ORF 02ep30607orf19 protein.
XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX OS Helicobacter pylori.
XX PN WO9737044-AL.
XX PD 09-OCT-1997.
XX PF 27-MAR-1997; 97WO-US05223.
XX PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX (ASTR) ASTRA AB.
XX PI Alm RA, Smith D;
XX DR WPI; 1997-503122/46.
DR N-PSDB; AAV24699.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX Claim 14; Page 524-525; 1145pp; English.
XX This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds. The
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 35679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in *E. coli*
 XX hosts.
 SQ Sequence 407 AA;

Query Match 26.9%; Score 1792.5; DB 18; Length 407;
 Best Local Similarity 89.2%; Pred. No. 4.9e-96;
 Matches 346; Conservative 17; Mismatches 18; Indels 7; Gaps 1;

QY 1 MEIOOHRKINRPLVSLALVGLVSIPTQOQSHAAFFTTVIIPAVGGTATGTVSGSL 60
 DB 17 meiqdthrklnrpvlsvlagallisaipqeshaaffttviiipavvgiatgtavtvs 76
 QY 61 LSWGLKQAEAEANKTPDKPKVRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGNAAR 120
 DB 77 lswglkqaeaeanktpdkpkvriqagkgfnefpnkeydlyrslsskidgwdgwnaar 136
 QY 121 HYWYKGGQONKLEVDMDKDAVGYTTLGLRNFTGGDLVNMQKATLRLGQFNGNSFTSYKD 180
 DB 137 hywvkggqwnklevdmdkdvgytklslrnfthgdlvnmqkatlrlgqfngnsftsykd 196
 QY 181 SADRTTRVDFNAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
 DB 197 sadrttrvfnaknisidnfeinrnrvsgagrkasstvtlqasegitssknaeislyd 256
 QY 241 GATNLASSSVKLMGNVWMLQVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
 DB 257 gatlnlasssvkmgvnmgrlqvgyalapsystintskvgevdfnhlvtvgdqnaga 316
 QY 301 GIANKKTNIGTLDLWQAGLNIIAPPEGGYKDKPNNTPSOGAKNDKQESS 360
 DB 317 gilasnkthigtldlwsaglniiappeggykdkpnstsgtkndkksqn----- 370
 QY 361 QNNSNTQVINPPNSAQKTEVQPTQVIDG 388
 DB 371 -nnsntevinnppntqkteteptksimg 397

RESULT 8
 AAW20652
 ID AAW20652 standard; protein; 407 AA.
 XX
 AC AAW20652;
 XX
 DT 14-JUL-1997 (first entry)
 XX
 H. pylori derived protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX
 OS Helicobacter pylori.
 XX
 PN W09640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 XX (ASTR) ASTRA AB.
 XX
 XX Berglindh OT, Smith D, Mellgaerd BL;
 XX WPI; 1997-052306/05.
 DR N-PSDB: AAT67905.
 XX
 XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent *H. pylori*

PT infection, and to detect *Helicobacter*
 XX Disclosure; Page 1074-1075; 1481pp; English.
 XX
 CC The present sequence is a *H. pylori* derived protein of unspecified
 CC function, no further details are given in the specification.
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*
 CC infection or to identify *H. pylori* polypeptide binding compounds,
 CC useful as potential *H. pylori* life cycle activators or inhibitors.
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely *H. pylori* antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide
 CC production, e.g. in *E. coli* hosts.
 XX
 SQ Sequence 407 AA;

Query Match 26.9%; Score 1792.5; DB 18; Length 407;
 Best Local Similarity 89.2%; Pred. No. 4.9e-96;
 Matches 346; Conservative 17; Mismatches 18; Indels 7; Gaps 1;

QY 1 MEIOOHRKINRPLVSLALVGLVSIPTQOQSHAAFFTTVIIPAVGGTATGTVSGSL 60
 DB 17 meiqdthrklnrpvlsvlagallisaipqeshaaffttviiipavvgiatgtavtvs 76
 QY 61 LSWGLKQAEAEANKTPDKPKVRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGNAAR 120
 DB 77 lswglkqaeaeanktpdkpkvriqagkgfnefpnkeydlyrslsskidgwdgwnaar 136
 QY 121 HYWYKGGQONKLEVDMDKDAVGYTTLGLRNFTGGDLVNMQKATLRLGQFNGNSFTSYKD 180
 DB 137 hywvkggqwnklevdmdkdvgytklslrnfthgdlvnmqkatlrlgqfngnsftsykd 196
 QY 181 SADRTTRVDFNAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
 DB 197 sadrttrvfnaknisidnfeinrnrvsgagrkasstvtlqasegitssknaeislyd 256
 QY 241 GATNLASSSVKLMGNVWMLQVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
 DB 257 gatlnlasssvkmgvnmgrlqvgyalapsystintskvgevdfnhlvtvgdqnaga 316
 QY 301 GIANKKTNIGTLDLWQAGLNIIAPPEGGYKDKPNNTPSOGAKNDKQESS 360
 DB 317 gilasnkthigtldlwsaglniiappeggykdkpnstsgtkndkksqn----- 370
 QY 361 QNNSNTQVINPPNSAQKTEVQPTQVIDG 388
 DB 371 -nnsntevinnppntqkteteptksimg 397

RESULT 9
 AAR79945
 ID AAR79945 standard; protein; 513 AA.
 XX
 AC AAR79945;
 XX
 DT 26-MAR-1996 (first entry)
 XX
 XX Helicobacter pylori antigenic protein.
 XX
 KW Vacuolating toxin; vaccine; immunisation; therapy; mutant;
 KW infection; *Helicobacter pylori*.
 XX
 OS Helicobacter pylori.
 XX
 PN W09522988-A1.
 XX

```
PD 31-AUG-1995.
XX
PF 23-FEB-1995; 95WO-US02219.
XX
PR 23-FEB-1994; 94US-0200232.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Blaser MJ, Cover TL;
XX
DR WPI: 1995-311383/40.
DR N-PSDB: AAT04133.
XX
PT Isolated DNA encoding Helicobacter pylori vacuolating toxin - useful
PT for immunisation against H. pylori infection
XX
PS Claim 19; Page 50-52; 64pp; English.
XX
CC Nucleic acid encoding the Helicobacter pylori vacuolating toxin
CC and a genetically altered mutant strain of H. pylori which contains
CC a foreign nucleic acid and does not express a functional vacuolating
CC toxin may be used to immunise a subject against H. pylori infection.
CC They may possibly also be used therapeutically. This antigenic
CC protein may also be used for immunisation purposes.
XX
SQ Sequence 513 AA;

Query Match 25.4%; Score 1694.5; DB 16; Length 513;
Best Local Similarity 63.3%; Pred. No. 3.2e-90;
Matches 343; Conservative 52; Mismatches 84; Indels 63; Gaps 7;

QY 311 GTLDLWQASAGLTIAPPEGYKDKPNTPSQGAKNDKSNESAKNDKQESSQNNSNTQVIN 370
Db 1 gvtldlwqasaglsitppegysktdkpsqspknd----- 37

QY 371 PPSNAQTEVQPTVIDGPFAGGKDTVNNINRINTNADGTVIRVGFKASLTNAAHLHG 430
Db 38 ----Lqkteiqptvldgpfaggkdtvnnifhlntkadgtlkaggfkaslttnaaahlhg 93

QY 431 KGGVNLNQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEKAGDTKNGT 490
Db 94 eggvnlnsqasgrslivenltgnitvegtlrvnnqvggaavagssanfekagadtnnat 153

QY 491 ATFNNDISLGRVNLKVDAAHT-----ANFKGIDT--GNGGFNT- 526
Db 154 atfnndihlgkavnrvdahtayfngniylgkstnlrvnghshahfnidatksdnglnts 213

QY 527 -LDFSGVTDKVNINKLITASTWAVKFNINELIVKNTNGISVGEYTHFSEDIGSQSRINT 585
Db 214 aldfsgvtdkvninklittsatvnnknfdikelvvttrvqsfqytfifgenigdksrigrv 273

QY 586 VRLETGTRSLFSGGVKFGKEGLVIDEFYSPWNYFDARNIKNVEITNKLAFGQGPSWG 645
Db 274 vrltgtrslfsggvkfgkeglvidefysspwnyfdarniknveitnklafgqgpswg 333

QY 646 TSKLMFNLLTQGNVMDYSQFSNLTIQGDFINNOGTINYLVRGKQVATLSVGNAAAMMF 705
Db 334 tsklmfnlltnosasmgykdlitigqhtnngatmlfvqdgrrvatlnaghaqsmif 393

QY 706 NNDIDSATGFYKPLKINSQAOLIKNTEHVLLKAKIIGYGNVSTGTNG-----ISNVN 758
Db 394 nndidsatgfykplkinnnaqltknkehlvlgkgrnidnlyv--gvqgasynaisasntn 451

QY 759 LEEQKERLALYNNNRMDTCVVR--NTDDIKACGMAIGDSVMNPNYKLYLIGKAWKN 816
Db 452 lleeqkerlalyannnrmdtcvvrntddikacgmaignsgmvpnnpndykyledkawkn 511

QY 817 IG 818
Db 512 tg 513
```

```
RESULT 10
AAB52500
ID AAB52500 standard; Protein; 326 AA.
XX
AC AAB52500;
XX
DT 23-FEB-2001 (first entry)
XX
DE Helicobacter pylori bait polypeptide #18.
XX
KW Helicobacter pylori; two-hybrid system; protein-protein interaction;
KW bait polypeptide; gastric ulcer; antibacterial.
XX
OS Helicobacter pylori.
XX
PN WO2000066722-A1.
XX
PD 09-NOV-2000.
XX
PF 14-APR-2000; 2000WO-IB00603.
XX
PR 30-APR-1999; 99EP-0401066.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Selig L, Rain J;
XX
DR WPI: 2000-687535/67.
DR N-PSDB: AAC97246.
XX
PT A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
PT cell clones -
XX
PS Example 5; Page 101; 267pp; English.
XX
CC The present sequence is a bait polypeptide used in a Helicobacter
CC pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX
SQ Sequence 326 AA;

Query Match 23.1%; Score 1538.5; DB 21; Length 326;
Best Local Similarity 90.3%; Pred. No. 1.9e-81;
Matches 299; Conservative 15; Mismatches 12; Indels 5; Gaps 1;

QY 207 VGSAGRKASSTVLTQLQASEGITSDKNAEISLYDGTATLNLASSVKLMGNVMGRQLYYVG 266
Db 1 vsgagrkasstvtlqasegitssknaeislydgtatlnlasnsvklmgvnmgrlqyyvg 60

QY 267 AYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQAQGIANKKTNIGTLDLWQSAGLNIAP 326
Db 61 aylapsystintskvtgevnfnhltvgdhnhaaqgiasnkhthigtldlwqsaglniap 120

QY 327 PEGGYKDKPNTPSQGAKNDKSNESAKNDKQESSQNNSNTQVINPPNSAQKTEVQPTQVI 386
Db 121 peggykdkpndtpsqgakndknesakndkqessqnnsntqvinnpsaqkteiqptqvi 175

QY 387 DGPFGAGKDTVNNINRINTNADGTVIRVGFKASLTNAAHLHGKGVNLNSQASGRSLI 446
Db 176 dgpfgagkdtvnnidrintnadtikvggkaslttnaaahlhgkgnlnsqasgrtll 235

QY 447 VENLTGNTIVDGPLRVNNQVGGYALAGSSANFEKAGTGTKNGTATFNNDISLGRVNLK 506
Db 1 venltgntivdgpplrvnnqvggalagssanfekagtgatfnndisigrvnlk 506
```


Qy	1009	SLSAEAVLYQAPKYEKPTNNWANAIGTSLNNGSNASLYGTSGAGVDAYLNGQVEAIVG	1068
Db	1	slesaeavlyqapkyekptnnwanaiggtslnsggnaslygtsgagvdaiylngeavlg	60
Qy	1069	GFGSGYSSFNPNRANSLNSCANNTNPGVYSRIFANQHEFDFAOCALGSDOSSLNFKSAL	1128
Db	61	gfsgysyssfnpnranslncanntnpgvysrifaqhefdfeagocalgsgdosslnfksal	120
Qy	1129	LQDLNQSRYHLAYSAATRASGYDFAFFRNALVKPSVGVSYNHLGSTNFKSNSTNOVAL	1188
Db	121	lrdlnqsyryhlaysaatrasgydydfaffrnalvlkpsvgvsnhlgstnfksnqkval	180
Qy	1189	KNGSSQSHLNFNASANYEARYYGDTSYFYMMNAGVYLQEFAHVGCSNNAASLNTFKYNAARNP	1248
Db	181	kngassqhlfnasanvearyyygdtisyfymnagvylqefanfsgsnavslntfkvnatrn	240
Qy	1249	LNTHARVMGGEGLKLAKEVFLNLGVYLLHNLISNIGHFASNLGHRYSF	1296
Db	241	lntharvmmggelklakevflnlglfvyllhnlisnighfasnlgmrysf	288
RESULT 12			
AAW55462	AAW55462 standard; Protein: 257 AA.		
XX	XX	AAW55462;	
XX	XX		
DT	24-JUN-1998	(first entry)	
XX	XX		
DE	H. pylori ORF 07ap11015_23938312_f3_2	cell envelope OMP.	
XX	XX		
KW	Cytoplasmic; vaccine; prevention; treatment; infection; envelope;		
KW	identification; binding compound; bacteria; life cycle; activator;		
KW	inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.		
XX	XX		
OS	Helicobacter pylori.		
XX	XX		
PN	WO9737044-A1.		
XX	XX		
PD	09-OCT-1997.		
XX	XX		
PF	27-MAR-1997; 97WO-US05223..		
XX	XX		
PR	06-DEC-1996; 96US-0761318.		
PR	29-MAR-1996; 96US-0625811.		
PR	02-APR-1996; 96US-0758731.		
PR	23-OCT-1996; 96US-0736905.		
PR	28-OCT-1996; 96US-0738859.		
XX	XX		
PA	(ASTR) ASTRA AB.		
PI	Alm RA, Smith D;		
XX	XX		
DR	WPI; 1997-503122/46.		
DR	N-PSDB; AAV24871.		
PT	Helicobacter pylori nucleic acid sequences and encoded		
PT	polypeptide(s) - useful in vaccines to treat or prevent H. pylori		
PT	infection and for diagnosis of H. pylori infection		
XX	XX		
PS	Disclosure; Page 667-668; 1145pp; English.		
XX	XX		
CC	This sequence is a H. pylori cell envelope outer membrane protein		
CC	(OMP) having a terminal Phe residue.		
CC	The protein may be used in a vaccine to prevent or treat H. pylori		
CC	infection or to identify H. pylori polypeptide binding compounds,		
CC	useful as potential H. pylori life cycle activators or inhibitors. The		
CC	DNA and probes derived from it may be used for the identification of		
CC	H. pylori in a sample and the diagnosis of H. pylori infection. Nuclei		
CC	acid sequences complementary to the DNA act as antisense sequences and		
CC	can be used to prevent the translation of H. pylori mRNA. Antibodies		
CC	against the protein can be used in immunoassays to evaluate the abunda		
CC	and distribution of H. pylori-specific antigens. The genomic sequence		

Search completed: August 16, 2002, 11:30:35
Job time: 191 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:27:24 ; Search time 18.86 Seconds
(without alignments)
1678.450 Million cell updates/sec

Title: US-09-921-157-3
Perfect score: 6671
Sequence: 1 MEIQQTHRKINRPLVSLV.....HNLISNIGHFASNLGMRYSF 1296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6671	100.0	1296	3	US-08-470-260-3	Sequence 3, Appli
2	6671	100.0	1296	3	US-08-471-491-3	Sequence 3, Appli
3	6671	100.0	1296	4	US-08-466-662-3	Sequence 3, Appli
4	6303.5	94.5	1287	1	US-08-200-232-2	Sequence 2, Appli
5	6303.5	94.5	1287	5	PCT-US95-02219-2	Sequence 2, Appli
6	6303.5	94.5	1287	5	PCT-US95-02219A-2	Sequence 2, Appli
7	1695.5	25.4	513	1	US-08-200-232-4	Sequence 4, Appli
8	1695.5	25.4	513	5	PCT-US95-02219-4	Sequence 4, Appli
9	1695.5	25.4	513	5	PCT-US95-02219A-4	Sequence 4, Appli
10	320.5	4.8	1612	1	US-08-169-927-2	Sequence 2, Appli
11	292	4.4	1529	2	US-08-728-470-10	Sequence 10, Appl
12	292	4.4	1529	4	US-08-719-641-10	Sequence 10, Appl
13	287.5	4.3	1338	2	US-08-728-470-9	Sequence 9, Appli
14	287.5	4.3	1338	4	US-08-719-641-9	Sequence 9, Appli
15	287.5	4.3	1599	2	US-08-617-697-9	Sequence 9, Appli
16	283	4.2	1541	4	US-08-296-791-3	Sequence 3, Appli
17	283	4.2	1541	5	PCT-US95-10661A-3	Sequence 3, Appli
18	276	4.1	1545	4	US-08-296-791-4	Sequence 4, Appli
19	276	4.1	1545	5	PCT-US95-10661A-4	Sequence 4, Appli
20	274	4.1	1600	2	US-08-617-697-10	Sequence 10, Appl
21	269	4.0	1536	1	US-08-038-682-2	Sequence 2, Appli
22	269	4.0	1536	1	US-08-302-832-2	Sequence 2, Appli
23	269	4.0	1536	2	US-08-530-198-2	Sequence 2, Appli
24	269	4.0	1536	2	US-08-469-880-2	Sequence 2, Appli
25	269	4.0	1536	2	US-08-728-470-2	Sequence 2, Appli
26	269	4.0	1536	2	US-08-617-697-2	Sequence 2, Appli
27	269	4.0	1536	4	US-08-719-641-2	Sequence 2, Appli

28	268	4.0	2048	4	US-09-268-347-48	Sequence 48, Appli
29	265.5	4.0	1702	4	US-08-296-791-5	Sequence 5, Appli
30	265.5	4.0	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
31	258.5	3.9	2314	4	US-09-268-347-49	Sequence 49, Appli
32	248	3.7	1477	1	US-08-038-682-4	Sequence 4, Appli
33	248	3.7	1477	1	US-08-302-832-4	Sequence 4, Appli
34	248	3.7	1477	2	US-08-530-198-4	Sequence 4, Appli
35	248	3.7	1477	2	US-08-469-880-4	Sequence 4, Appli
36	248	3.7	1477	2	US-08-728-470-4	Sequence 4, Appli
37	248	3.7	1477	2	US-08-617-697-4	Sequence 4, Appli
38	248	3.7	1477	4	US-08-719-641-4	Sequence 4, Appli
39	244	3.7	2353	4	US-09-377-155-33	Sequence 33, Appli
40	244	3.7	2353	4	US-08-913-942-4	Sequence 4, Appli
41	244	3.7	2353	4	US-09-669-974-33	Sequence 33, Appli
42	243	3.6	1912	1	US-08-409-995-4	Sequence 4, Appli
43	243	3.6	1912	3	US-08-685-467-4	Sequence 4, Appli
44	242	3.6	2354	4	US-09-268-347-47	Sequence 47, Appli
45	238	3.6	1394	4	US-08-296-791-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-470-260-3
; Sequence 3, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-260-3

Query Match 100.0%; Score 6671; DB 3; Length 1296;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1296;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	MEIQOTHRKNRPLVSLALVGLVSIITPQOSHAAFFFTVIIIPAIVGGIATGTAVGTVSGL	60
DB	1	MEIQOTHRKNRPLVSLALVGLVSIITPQOSHAAFFFTVIIIPAIVGGIATGTAVGTVSGL	60
QY	61	LSWGLKQAEANKTPDPKDKWRTOAQKGFNEFPNKEDYLRSLSSKIDGWDWGNAAAR	120
DB	61	LSWGLKQAEANKTPDPKDKWYRIQAGKGFNEFPNKEDYLRSLSSKIDGWDWGNAAAR	120
QY	121	HYWYKGOQNKLEVDMKDAVGTYYTSLGRNFTGGDLDNVMOKATILRIGQFNGNSFTSYKD	180
DB	121	HYWYKGOQNKLEVDMKDAVGTYYTSLGRNFTGGDLDNVMOKATILRIGQFNGNSFTSYKD	180
QY	181	SADRTTRVDFNAKNISIDNFVEINNRVSGSAGRKASSTVLTLQASEGITSDKNAEISLYD	240
DB	181	SADRTTRVDFNAKNISIDNFVEINNRVSGSAGRKASSTVLTLQASEGITSDKNAEISLYD	240
QY	241	GATLNLASSSVKLMGNVMMGLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA	300
DB	241	GATLNLASSSVKLMGNVMMGLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA	300
QY	301	GIANKKNTIGTLDWQASAGLNIIAPPEGGYKDKPNNTPSOGAKNDKNSAKNDKQESS	360
DB	301	GIANKKNTIGTLDWQASAGLNIIAPPEGGYKDKPNNTPSOGAKNDKNSAKNDKQESS	360
QY	361	QNSNTQVIPPNSAQKTEVQPTQIDGPPAGGKDTVVNINRIINTADGTTIRVGGFKASL	420
DB	361	QNSNTQVIPPNSAQKTEVQPTQIDGPPAGGKDTVVNINRIINTADGTTIRVGGFKASL	420
QY	421	TTNAHLHIKGGVNLNSQASGRSLIVENLTGNTITVDCGPLRVNNQVGGYALAGSSANPEF	480
DB	421	TTNAHLHIKGGVNLNSQASGRSLIVENLTGNTITVDCGPLRVNNQVGGYALAGSSANPEF	480
QY	481	KAGPDTKNGATFNNDISLGRFVNLKVDAHTANPKGIDTNGGNTLDFSGVTDKVNINK	540
DB	481	KAGPDTKNGATFNNDISLGRFVNLKVDAHTANPKGIDTNGGNTLDFSGVTDKVNINK	540
QY	541	LITASTNVAVKFNINELIVKNTGISVGEYTHFSEDIGSQSRINTVRLTETSRLSFSGGV	600
DB	541	LITASTNVAVKFNINELIVKNTGISVGEYTHFSEDIGSQSRINTVRLTETSRLSFSGGV	600
QY	601	KFKGGEKLVIDEFYYPSPWYFDARNIKVETNKLAFPGQSPNGTSKLMMFNNTLQONA	660
DB	601	KFKGGEKLVIDEFYYPSPWYFDARNIKVETNKLAFPGQSPNGTSKLMMFNNTLQONA	660
QY	661	VMDYSQFSNLTIOGDFTNNGQTNLYLRGGKVAATLSVGNAAAMFNNDIDSATGFYKPLI	720
DB	661	VMDYSQFSNLTIOGDFTNNGQTNLYLRGGKVAATLSVGNAAAMFNNDIDSATGFYKPLI	720
QY	721	KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEBQFKERLALYNNNRMDTCV	780
DB	721	KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEBQFKERLALYNNNRMDTCV	780
QY	781	VRNTDDIKACGMALGDOSWVNNPNYKYLIGKAKNKGISKTANGSKISVYILGNSTPTE	840
DB	781	VRNTDDIKACGMALGDOSWVNNPNYKYLIGKAKNKGISKTANGSKISVYILGNSTPTE	840
QY	841	NGGNTNLPTNTTNSAKSANNALQAONPAFQPSATPNLVAINQHDFTGIESVFELANRSK	900
DB	841	NGGNTNLPTNTTNSAKSANNALQAONPAFQPSATPNLVAINQHDFTGIESVFELANRSK	900
QY	901	DIDTLYNSGAQGRDLQLTLLIDSHDAGYARKMTIDATSANEITKOLNTATTTLNNIASLE	960
DB	901	DIDTLYNSGAQGRDLQLTLLIDSHDAGYARKMTIDATSANEITKOLNTATTTLNNIASLE	960
QY	961	HKTSGLQTLTSLSNAMILNSRLVNLRSRHTNHIIDSFARKLQALKQOKFASLESAAEVLYQF	1020
DB	961	HKTSGLQTLTSLSNAMILNSRLVNLRSRHTNHIIDSFARKLQALKQOKFASLESAAEVLYQF	1020
QY	1021	APKYKEFTNVWANAIGGTSNLNGNSNASTLYGTSAGVDAYLNGQVPAIVGGFGSYSSFN	1080
DB	1021	APKYKEFTNVWANAIGGTSNLNGNSNASTLYGTSAGVDAYLNGQVPAIVGGFGSYSSFN	1080

D	b	1021	APKYEKPTNVWANAIGGTSNLNGSNASLIYGTSGAVDAVLCNQVEAIVCGFSGSYGYSSFN
Q	y	1081	RANSLNSGANTNFGVYSRIFANQHEFDFAQAAGALGSDQSLSLNFKSALLQDLINQSHYLA 11
D	b	1081	RANSLNSGANTNFGVYSRIFANQHEFDFAQAAGALGSDQSLSLNFKSALLQDLINQSHYLA 1140
Q	y	1141	YSAATRASYGDYDFAFRNALVLKPSVGYSYNHLGSTNFKSNSTNQVALKNSSSQHLENA 1200
D	b	1141	YSAATRASYGDYDFAFRNALVLKPSVGYSYNHLGSTNFKSNSTNQVALKNSSSQHLENA 1200
Q	y	1201	SANYEARYYGDTSYFYMNAGVLOEFAHVGSNNAASSLATFKVNAARNPLNTHARVMGMGE 1260
D	b	1201	SANYEARYYGDTSYFYMNAGVLOEFAHVGSNNAASSLATFKVNAARNPLNTHARVMGMGE 1260
Q	y	1261	LKLAKEVELNLGVVYLHNLLISNIHGHSFASNLGMRYSF 1296
D	b	1261	LKLAKEVELNLGVVYLHNLLISNIHGHSFASNLGMRYSF 1296
 RESULT 2 US-08-471-491-3 ; Sequence 3, Application US/08471491B ; Patent No. 6090611 ; GENERAL INFORMATION: ; APPLICANT: Covacci, Antonello ; APPLICANT: Bugnoli, Massimo ; APPLICANT: Telford, John ; APPLICANT: Macchia, Giovanni ; APPLICANT: Rappuoli, Rino ; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And ; FILE REFERENCE: CHIR0044 ; CURRENT APPLICATION NUMBER: US/08/471.491B ; CURRENT FILING DATE: 1995-06-06 ; NUMBER OF SEQ ID NOS: 8 ; SOFTWARE: Patent In Ver. 2.1 ; SEQ ID NO 3 ; LENGTH: 1296 ; TYPE: PR1 ; ORGANISM: Helicobacter pylori US-08-471-491-3			
 Query Match 100.0%; Score 6671; DB 3; Length 1296; Best Local Similarity 100.0%; Pred. No. 0; Matches 1296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Q	y	1	MEIOOTHRKINRPLVSALVGLYSITPQQSHAAFFTTVIIIPAIVGGTATGTA VGTVSGL 60
D	b	1	MEIOOTHRKINRPLVSALVGLYSITPQQSHAAFFTTVIIIPAIVGGTATGTA VGTVSGL 60
Q	y	61	LSWGLKQAEANKTPDKPKVWRIOAQKGFEFPNKEDLYRSLLSSKIDGWDWGNAAR 120
D	b	61	LSWGLKQAEANKTPDKPKVWRIOAQKGFEFPNKEDLYRSLLSSKIDGWDWGNAAR 120
Q	y	121	HYWKGGQNKLVDMDKDAVTYTILSGLURNFTGGDLDVNMOKATRLRGQFNNGSFTSYKD 180
D	b	121	HYWKGGQNKLVDMDKDAVTYTILSGLURNFTGGDLDVNMOKATRLRGQFNNGSFTSYKD 180
Q	y	181	SADRTTRVDFNNAKNISIDNFVEINNRRVSGAGRASSVTLTLQASEGTTSDKNAEISLYD 240
D	b	181	SADRTTRVDFNNAKNISIDNFVEINNRRVSGAGRASSVTLTLQASEGTTSDKNAEISLYD 240
Q	y	241	GATLNLA SSSVKLMGNVMWGRQLQVYGAYLAPYSTINTSKVTGEVNFNHLTVGDKNA AQA 300
D	b	241	GATLNLA SSSVKLMGNVMWGRQLQVYGAYLAPYSTINTSKVTGEVNFNHLTVGDKNA AQA 300
Q	y	301	GIANKKTNICLTLDLWQAGLNIITAPPBGKYDKPNNTPSOGAKNDKNE SAKNDKQESS 360
D	b	301	GIANKKTNICLTLDLWQAGLNIITAPPBGKYDKPNNTPSOGAKNDKNE SAKNDKQESS 360
Q	y	361	QNNSTQVINPPNSAQKTEVPQTVIDGPFA GKKDTVVYINRINTNADGTRVGGFKASL 420

Qy 841 NGGNTNLPTNTTNSARSANNAALQAPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
Db 841 NGGNTNLPTNTTNSARSANNAALQAPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
Qy 901 DIDTLYANSQAQGRDLLQTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
Db 901 DIDTLYANSQAQGRDLLQTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
Qy 961 HKTSGLOTLSLSNAMILNSRLVNLRSRHTNHIDSFARKLQALKDQKFALESAAEVLVYQF 1020
Db 961 HKTSGLOTLSLSNAMILNSRLVNLRSRHTNHIDSFARKLQALKDQKFALESAAEVLVYQF 1020
Qy 1021 APKYKPTNWMANAIGGTSNLNGSNASLYGTSAGVAYLNGQVEAIVGGFGSYGSFNN 1080
Db 1021 APKYKPTNWMANAIGGTSNLNGSNASLYGTSAGVAYLNGQVEAIVGGFGSYGSFNN 1080
Qy 1081 RANSLNSGANTNFVYSRIFANQHEFDFAQAGALGSDQSSLNFKSALLQDLNQSYYHLA 1140
Db 1081 RANSLNSGANTNFVYSRIFANQHEFDFAQAGALGSDQSSLNFKSALLQDLNQSYYHLA 1140
Qy 1141 YSAATRASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSQHLFNA 1200
Db 1141 YSAATRASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSQHLFNA 1200
Qy 1201 SANYEARYYYGDTSYFYMNAGVLOQFAHVGSGNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Db 1201 SANYEARYYYGDTSYFYMNAGVLOQFAHVGSGNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Qy 1261 LKLAKEVFLNLGVVYLNHLISNIGHFASNLGMRYSF 1296
Db 1261 LKLAKEVFLNLGVVYLNHLISNIGHFASNLGMRYSF 1296

RESULT 4

US-08-200-232-2
; Sequence 2, Application US/08200232
; Patent No. 5721349
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; APPLICANT: Blaser, Martin J.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND RELATED METHODS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,232
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-200-232-2

Query Match 94.5%; Score 6303.5; DB 1; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;
Qy 1 MEIQOTHRKNRPLVSLALVGLVSIITPQQSHAAFFTTVIIIPAIIVGGIATGATVGTVSGL 60
Db 1 MEIQOTHRKNRPLVSLALVGLVSIITPQQSHAAFFTTVIIIPAIIVGGIATGATVGTVSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVRIQACKGFNEFPNKEYDLYRSLSSKIDCGWDGWNAR 120
Db 61 LSWGLKQAEANKTPDKPKVRIQACKGFNEFPNKEYDLYRSLSSKIDCGWDGWNAR 120
Qy 121 HYWVGKQONKLEVDMDKAVGTYTLISGLRNFPGDLDVNMOKATLRLGQFNGNSFTSYKD 180
Db 121 HYWVGKQONKLEVDMDKAVGTYTLISGLRNFPGDLDVNMOKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDENAKNISIDNFEVNNRVGSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
Db 181 SADRTTRVDENAKNILIDNFEVNNRVGSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWMLQVYGLAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATNLASSSVKLMGNVWMLQVYGLAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Qy 301 GIIANKNTNIGTLDWOSAGLNIIPPEGGYKDKPNNTPSQSGAKNDKSKDKQESS 360
Db 301 GIIANKNTNIGTLDWOSAGLNIIPPEGGYKDKPNNTPSQSGAKNDKSKDKQESS 360
Qy 361 QNNSNTQVIPPNSAOKTEVQPTQVIDGPFAGGKDTVVYNNRINTNADGTIRVGGFKASL 420
Db 361 QNNSNTQVIPPNSAOKTEVQPTQVIDGPFAGGKDTVVYNNRINTNADGTIRVGGFKASL 420
Qy 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGGYALAGSANPEF 480
Db 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGGYALAGSANPEF 480
Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDATANKGIDTGNNGENTLDFSGVTOKVINK 540
Db 481 KAGDTKNGTATFNNDISLGRFVNLKVDATANKGIDTGNNGENTLDFSGVTOKVINK 540
Qy 541 LITASTNVAVKFNINELIVKTNISVGEYTHFSEIDIGSOSRINTVRLTGTSLFSGGV 600
Db 541 LITASTNVAVKFNINELIVKTNISVGEYTHFSEIDIGSOSRINTVRLTGTSLFSGGV 600
Qy 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPOGSPWGTSKLMFNNTLQNA 660
Db 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPOGSPWGTSKLMFNNTLQNA 660
Qy 661 VMDYSQFSLTIQGFINNOCITINYLVRGKQVATLSVGNAAAMFNNDIDSATGYKPLI 720
Db 661 VMDYSQFSLTIQGFINNOCITINYLVRGKQVATLSVGNAAAMFNNDIDSATGYKPLI 720
Qy 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 780
Qy 781 VRNTDDIKACGMAIGDQSWNNPNYKYLICKANKNTGISKTANGSKISVYLLNSPTPE 840
Db 781 VRNTDDIKACGMAIGDQSWNNPNYKYLICKANKNTGISKTANGSKISVYLLNSPTPE 840
Qy 841 NGGNTNLPTNTTNSARSANNAALQAPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
Db 841 NGGNTNLPTNTTNSARSANNAALQAPFAQPSATPNLVAIQHDFGCTIESVFELANRSK 900
Qy 901 DIDTLYANSQAQGRDLLQTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
Db 901 DIDTLYANSQAQGRDLLQTLIDSHDAGYARKMIDATSAEITKOLNATATTLNIIASLE 960
Qy 961 HKTSGLOTLSLSNAMILNSRLVNLRSRHTNHIDSFARKLQALKDQKFALESAAEVLVYQF 1020
Db 961 HKTSGLOTLSLSNAMILNSRLVNLRSRHTNHIDSFARKLQALKDQKFALESAAEVLVYQF 1020

Db 952 HKTSSLQTLSLSNAMILNRLVNLRRHTNNIDSFARLQALQDQRFASLESAAEVLQF 1011
QY 1021 APYKPTNWNANAIGTSLNNGSNASLYGTSGAYDAYLNGQVEAIVGGFGSYGYSFNN 1080
Db 1012 APYKPTNWNANAIGASLNGGNASLYGTSGAYDAYLNGQVEAIVGGFGSYGYSFNN 1071
QY 1081 RANSLNGANNNTNFGVYSRIFANQHEFDEFAQAGALGSDQSSLNFKSALLQDLNQSYYLA 1140
Db 1072 QANSLNGANNNTNFGVYSRIFANQHEFDEFAQAGALGSDQSSLNFKSALLQDLNQSYYLA 1131
QY 1141 YSAATRASGYDFAFRNALVLKPSVGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1132 YSAATRASGYDFAFRNALVLKPSVGVSYNHLGSTNFKSNSTNFKVALNKGSSQHLFNA 1191
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGE 1260
Db 1192 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGE 1251
QY 1261 LKLAKEVFLNLGVVYVYHNLISNIGHGFAASNLGMRYSF 1296
Db 1252 LKLAKEVFLNLGVVYVYHNLISNIGHGFAASNLGMRYSF 1287

RESULT 5
PCT-US95-02219-2
; Sequence 2, Application PC/TUS9502219
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND RELATED METHODS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02219
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-02219-2

Query Match 94.5%; Score 6303.5; DB 5; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;

QY 1 MEIQOQTHRKINRPLVSLALVGALVSTTPQOSHAAFTTIVIIPAIVGGIATGTAVGTVSG 60
Db 1 MEIQOQTHRKINRPLVSLALVGALVSTTPQOSHAAFTTIVIIPAIVGGIATGTAVGTVSG 60
QY 61 LSNGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120

Db 61 LGWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAT 120
QY 121 HWYKVGQGNKLEVDKMDAVGYITLSGLNFTGGDLVDNMQKATLRLGFGNFSFTSYKD 180
Db 121 HWYKVGQGNKLEVDKMDAVGYITLSGLNFTGGDLVDNMQKATLRLGFGNFSFTSYKD 180
QY 181 SADRTTRVDFNAKNTSIDNFVEINNRVSGAGRKASSTVLTLOASBGITSDKNAEISLYD 240
Db 181 SADRTTRVDFNAKNTSIDNFVEINNRVSGAGRKASSTVLTLOASBGITSSKNAEISLYD 240
QY 241 GATLNLASSSVKLMGNWVMGRLOYVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
Db 241 GATLNLASSSVKLMGNWVMGRLOYVGYLAPSYSTINTSKVTGEVNFNHLTVGDKNAAQA 300
QY 301 GIITANKKTNTGTLWQSLNIIAPPEGGYKDKPNPTPSQCAKNDKNESAKNDKQESS 360
Db 301 GIITANKKTNTGTLWQSLNIIAPPEGGYKDKPNPTPSQSG-----AKNDKQESS 352
QY 361 QNNSNTQVINPPNSAOKTEVQPTQVIDGPFAGGKDFVWVNIINRINTNADGTIRVGGFKASL 420
Db 353 QNNSNTQVINPPNSTQKTQTEVQPTQVIDGPFAGGKDFVWVNIINRINTKADGTIKVGGFKASL 412
QY 421 TTNAHLHIGKGVNLSNOASGRSLIVENLTGNITVDGPLRVNNOVGGYALAGSSANFEF 480
Db 413 TTNAHLHIGKGVNLSNOASGRSLIVENLTGNITVDGPLRVNNOVGGYALAGSSANFEF 472
QY 481 KAGTDTKNGTATFNNDISLGRFVNLKVDATNFAGKIDTGNGGFTLDFSGVTDKVNINK 540
Db 473 KAGVDTKNGTATFNNDISLGRFVNLKVDATNFAGKIDTGNGGFTLDFSGVTNKVNINK 532
QY 541 LITASTNVAVKNFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSLPSGGV 600
Db 533 LITASTNVAVKNFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSLPSGGV 592
QY 601 KFKGSEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPGWGTSKLMFNLLTGQNA 660
Db 593 KFKGSEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPGWGTSKLMFNLLTGQNA 652
QY 661 VMDYSQFSLTTOGDFINNOGTINYLVRGKVTATLSVGNAAAMFNNDISATGFKPLI 720
Db 653 VMDYSQFSLTTOGDFINNOGTINYLVRGKVTATLSVGNAAAMFNNDISATGFKPLI 712
QY 721 KINSADLLIKNTEHVLKAKIIGYGNVSTGNGISVNLLEEQFKERLALYNNNRMDTCV 780
Db 713 KINSADLLIKNTEHVLKAKIIGYGNVSTGNGISVNLLEEQFKERLALYNNNRMDTCV 772
QY 781 VRNTDDIKACGMAIGDQSMVNNPDNYKYLIQKAWKNIGISKTAGSKISVYVILGNSTPTE 840
Db 773 VRNTDDIKACGMAIGDQSMVNNPDNYKYLIQKAWKNIGISKTAGSKISVYVILGNSTPTE 832
QY 841 NGGNTNLTPTNTSNARSANALAQNAQPAQPSATPNLVAIINQHDGTTIESVFELANRSK 900
Db 833 NGGNTNLTPTNTNNARFASYALIKNAPFAH-SATPNLVAIINQHDGTTIESVFELANRSK 891
QY 901 DIDTLYANSQAQGRLLQTLTLLIDSDHAGYARKMIDATSAEITKOLNTATTTLNINIASLE 960
Db 892 DIDTLYANSQAQGRLLQTLTLLIDSDHAGYARKMIDATSAEITKOLNTATTTLNINIASLE 951
QY 961 HKTSSLQTLSLSNAMILNRLVNLRRHTNNIDSFARLQALQDQRFASLESAAEVLQF 1020
Db 952 HKTSSLQTLSLSNAMILNRLVNLRRHTNNIDSFARLQALQDQRFASLESAAEVLQF 1011
QY 1021 APYKPTNWNANAIGTSLNNGSNASLYGTSGAYDAYLNGQVEAIVGGFGSYGYSFNN 1080
Db 1012 APYKPTNWNANAIGASLNGGNASLYGTSGAYDAYLNGQVEAIVGGFGSYGYSFNN 1071
QY 1081 RANSLNGANNNTNFGVYSRIFANQHEFDEFAQAGALGSDQSSLNFKSALLQDLNQSYYLA 1140
Db 1072 QANSLNGANNNTNFGVYSRIFANQHEFDEFAQAGALGSDQSSLNFKSALLQDLNQSYYLA 1131
QY 1141 YSAATRASGYDFAFRNALVLKPSVGVSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200

Db	1132	YSAATRASVGYDEAFERNALVLKPSGVGSYVNHLGSTNFKNSTNKKVALSNGSSSQHLPNA	1191
Qy	1201	SANVEARYYYGDTSEYPMNAGVLQSEFAHVGSNNAAASLNTFKVNAARNPLNTHARVMGGE	1260
Db	1192	SANVEARYYYGDTSEYPMNAGVLQSEFAHVGSNNAAASLNTFKVNAARNPLNTHARVMGGE	1251
Qy	1261	LKLAKEVFNGLGVVYLNRNLSINIGHFASNLGMRYSF	1296
Db	1252	LKLAKEVFNGLGVVYLNRNLSINIGHFASNLGMRYSF	1287

```

RESULT      6
PCI-US95-02219A-2
; Sequence 2, Application PC/TUS9502219A
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; APPLICANT: Tummur, Murali KR
; APPLICANT: Cao, Ping
; APPLICANT: Thompson, Stuart A.
; APPLICANT: Blaser, Martin J.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND THE RELATED METHODS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02219A

```

Query Match	94.5%	Score 6303.5;	DB 5;	Length 1287;
Best Local Similarity	94.7%;	Pred. No. 0;		
Matches 1227;	Conservative	28;	Mismatches	32;
			Indels	9;
			Gaps	2;

Qy	1	MEIQOTHRKNRPLVSLALVGLVSTTPQOSHAAFFTIIIPALVGGIATGTAVGTVSGL	60
Db	1	MEIQOTHRKNRPLVSLALVGLVSTTPQOSHAAFFTIIIPALVGGIATGTAVGTVSGL	60
Qy	61	LSWGLQAEANKTPDKPKVRIQAGKGFNEPPNKEYDLYRSLSSKIDGGWDMGNAAR	120
Db	61	LGWGLQAEANKTPDKPKVRIQAGKGFNEPPNKEYDLYKSLSSKIDGGWDMGNAAT	120
Qy	121	HYWYKGGQONKLEVDMMKDAVGTVTLSGLRNFTGGDLVDNMQKATRLGQFNGNSFTSYKD	180
Db	121	HYWIKGGQWNKLEVDMMKDAVGTVKLSGLRNFTGGDLVDNMQKATRLGQFNGNSFTSYKD	180
Qy	181	SAERTTRVDFNAKNISIDNPEVINNRVGGAGKRSSTVLTQASBGITSDKNNAEISLYD	240

```
RESULT 7
US-08-200-232-4
; Sequence 4, Application US/08200232
; Patent No. 5721349
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; APPLICANT: Blaser, Martin J.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND RELATED METHODS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200.232
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-232-4

Query Match 25.4%; Score 1695.5; DB 1; Length 513;
Best Local Similarity 63.3%; Pred. No. 8.8e-107;
Matches 343; Conservative 52; Mismatches 84; Indels 63; Gaps 7;

QY 311 GTDLWQAGLNIAPPEGGYKDKPNTPSQSGAKNDKSNKAKNDKQESSQNNSNTQVIN 370
1 GVLDLWQAGLSIIIPPEGGYESKTKDNPSONSPKND----- 37
QY 371 PPSNAQKTEVQPTQIDGPFAGGKDTVYVNIININTNADGTIRVGGFKASLTNAHLHIG 430
DB 38 ----TQKTEIQPTQIDGPFAGGKDTVYVNIININTNADGTIRVGGFKASLTNAHLHIG 93
QY 431 KGGVNLNOASGRSLIVENLTGNITVDGFLRVNNQVGGYALAGSSANPEFKAGTDTKNT 490
DB 94 EGGVNLNOASGRSLIVENLTGNITVDGFLRVNNQVGGYALAGSSANPEFKAGTDTKNT 490
QY 491 ATFNNDISLGRFVNLKVDAT-----ANFKGIDT--GNGGFNT- 526
DB 154 ATFNNDIHLGRAVNLRVDAHTATYENGNTYLGKSTNLRVNGHSAHEKNIDATKSDNGLNLS 213
QY 527 -LDFSGVTDKVNINKLITASTVAVAKNENINELIVTKNCISVGEYTHSESDIGSQSRINT 585
DB 214 ALDFSGVTDKVNINKLITASTVAVAKNENINELIVTKNCISVGEYTHSESDIGSQSRIGV 273
QY 586 VRLFTGTRSLRSGGKFKGGEKLVIDEFYISPNWYFDARNIKNVEITNKLAFPGQSPWG 645
DB 274 VSLQTGYSPAYSGGVTFKSGKLVIDEIYHAPWYFDARNVTDVVEINKRILFCAPGNLAG 333
QY 646 TSKLMFNNTLGNQAVMDYSQFSNLTIOGDFINNOGCTINLVVRGKGVATLSYGNAAAMMF 705

Db 334 KTGLMFNLTLSNASMDYKDLDTIOGFTNNOCTMNLFFVQDGRVATLNAGHOASMF 393
QY 706 NNDIDSATGYFKPLIKINSAODLIKNTHEVLLKAKIIGYGNVSTGTNG-----ISNVN 758
DB 394 NNLVDSATGYFKPLIKINNAOQLTRKHEVLVKGKRNIDNLY--GVQGASYDNISASNTN 451
QY 759 LEEQFKERLALYNNNNRMDTCVVR--NTDDIKACGMAIGDQSWNNPNDYKYLIGKAWKN 816
DB 452 LOEQFKERLALYNNNNRMDICVVRKGNITDDIKACGMAIGNQSWNNPNDYKYLEGKAWN 511
QY 817 IG 818
DB 512 TG 513

RESULT 8
PCT-US95-02219-4
; Sequence 4, Application PC/TUS9502219
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; APPLICANT: Blaser, Martin J.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND RELATED METHODS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02219
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-02219-4

Query Match 25.4%; Score 1695.5; DB 5; Length 513;
Best Local Similarity 63.3%; Pred. No. 8.8e-107;
Matches 343; Conservative 52; Mismatches 84; Indels 63; Gaps 7;

QY 311 GTDLWQAGLNIAPPEGGYKDKPNTPSQSGAKNDKSNKAKNDKQESSQNNSNTQVIN 370
1 GVLDLWQAGLSIIIPPEGGYESKTKDNPSONSPKND----- 37
QY 371 PPSNAQKTEVQPTQIDGPFAGGKDTVYVNIININTNADGTIRVGGFKASLTNAHLHIG 430
DB 38 ----TQKTEIQPTQIDGPFAGGKDTVYVNIININTNADGTIRVGGFKASLTNAHLHIG 93
QY 431 KGGVNLNOASGRSLIVENLTGNITVDGFLRVNNQVGGYALAGSSANPEFKAGTDTKNT 490
DB 94 EGGVNLNOASGRSLIVENLTGNITVDGFLRVNNQVGGYALAGSSANPEFKAGTDTKNT 490
```


CLASSIFICATION: 435
PRIOR APPLICATION DATA: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-927-2

Query Match 4.8%; Score 320.5; DB 1; Length 1612;
Best Local Similarity 18.7%; Pred. No. 3.8e-13;
Matches 293; Conservative 186; Mismatches 555; Indels 533; Gaps 69;

QY 20 VGLVSIPTQSSHAFFTVIIPAVGGIATGTAFTVSGLLSGWGLKQAEAEANKTPDKPD 79
DB 77 LGFTITIAQO---AKFFNFV-----AAGKILNITGGI-TVQEAASNTINAQN 120
QY 80 KVRIRIAGKGFN-----EPP-----NKEYDYLRLSSLLSSKIDGGDW 115
DB 121 ALTKVHGGAANDLSGLSITFAVCPVLEFNLINPINSRSSYHLVSNRSKI----- 174
QY 116 GNAARHWYKGGQONKLEVDMDKDAVGYTLGLGRNFTGGDLVNMOK-----ATLRLGQFN 171
DB 175 -----VNGG-----NGILNTNGFIQVSDNTFAGIKTINIDDCQ 208
QY 172 GNSFTYKDSADTRVRDNKAKNISI-DNFVEINRVGSG-----AGRKASSTVLTLOAS- 225
DB 209 GLMFNSTPDA-----NTLQAGGNTINENGIDGTGKLVLSKNGAATFNFVTGTL 260
QY 226 ---EGITSKDNAEISLYDGATLNLASSVKLMG-----NVWGRLOQVVGAY 268
DB 261 GGNLKGIIELNIAA---GKLISLGGAAVAVGTONGAGRAAGFIVSDVNGNAATISGQ 317
QY 269 LAPSYTINTSKYTGVEFNHL-----TVGDKNAQAAGIIANK-----KTNIGTLDLWQ 317
DB 318 VYAKNVVIOGANNAGQVTFEHIIVDVGLGCTTNEKTADSKVIITENSFGSTFNGNLD--- 374
QY 318 SAGLNIAPP-----EGYKDKPNTPSQG-----AKNDKNESAKNDKQESSQ 361
DB 375 ---TQIVVPDKILGNFIDGVKNNGNTAGVITFNANGALVSASTDPNIAVTNINAIEAE 431
QY 362 NNSNTQV-----INPPNSAQTEVQPTQVDPGAGGKDTVVNINRINTN---ADGTI 411
DB 432 GAGVVELSGIHIAELRLNGSGSTFKLADGTGVINGP-----VNQALMNNNALAAGSI 483
QY 412 RVGGFKASLTNAHLHIGKGVN-----LSNQAS-----GRSLIVENLTGNITVD- 457
DB 484 QLDG-SAITGD-----ICNGVYNAALQHTITLANDASKILALDGANIGANVGGAITHFQA 537
QY 458 --GPLRVNNOVGVALAGSANEFFKAGTD-----TKNGTATFN----- 494
DB 538 NGGTIKLTNTQNNIV-----NFDLDTTDTKTVGDVASSLTNTQTLTNGISIGTVVANTK 592
QY 495 -----NDISLGRV---NLKVDHAITANF---KGIDTNGCGFNTLDFSGV 532
DB 593 TLAQLNIGSKTILNAGDVAINELVNIENSGVOLNHNNTYLIITKTINAANOG-----QIIVA 648
QY 533 TDKVNTNKLITATNV-----AVKFNENELI-----VKNTGISVSG 568
DB 649 ADPLNTNTIADGTNLGSAENPLSTIHFATKAANADSLNVGKVNLYANNITNDANVG 708
QY 569 EYTHF-----SEDIGSQS---RINTVRLETGTSLFSGGVFKGGEKLIVDEFYSPW 618

DB 709 S-LHFRSGGTSIVSGTVGGQGHKLNLLDNGCTTVKFLGDTTFNGGTKIEGKSILQISN 767
QY 619 NYF-----DARNIKNVEITN---KLAFGPGQSPWGTSK-----LMFN---NLTLG 657
DB 768 NYTTDHVESADNTGTLEFVNTDPTVTLNKGAYFVLKQVLIISGGPGNIVFNEIGNVGIV 827
QY 658 QNAVMDYSQFSLTIQGFINNQTINYLVRGKGKATLSVGNAAAMFNNDIDSATGCFYK 717
DB 828 HGTAANSISFENASL-----GTSIFLPSGTPDLVLTIKSVG---NGVDN---FNA 873
QY 718 PLIKINSADOLIKNTEHVLKAKIIG-----YGNVSTGTNGISNVNLEE 761
DB 874 PIVVSGIDSMINNGQIIGDKNIIALSGLSDNSITVNANTLYSGIRTTKNNQGTVTLSG 933
QY 762 QPKER-----LALYNNNRMD---TCVVRNTDDIKACGMAIGDQ----- 797
DB 934 GMPNPCTIYGLGLENGSPKLGKQVTFDTDYNNGLSIANNVTINDDVTLTGTGIAGTDFD 993
QY 798 -----SWVANNPDNYKYLIGKAWKNIGISKITANGSKISVYVLGNS-----TPTEN-- 841
DB 994 AKITLGSVNGNANVREVDSTFSDPRSMIVATQANKGTIVYLGNALYSNIGSLDTPVASVR 1053
QY 842 ---NLPTNTNARSANNALQAONAPPAOPSAT-----CGNTT----- 846
DB 1054 FTGNDSCAGLOGNIYSONIDFGTYNLTLNSNVILGGTTAINGEIDLTLNLIIFANGTS 1113
QY 847 ---NLPTNTNARSANNALQAONAPPAOPSAT-----CGNTT----- 846
DB 1114 TWGDNSTISITLNVSSNGIGQVVIADQAVNATTTGTTIKIODNANANFSGQAVTLIQ 1173
QY 876 -----PNLVAINOHDGFTIESVFE-----LANRSKD-----IDTLYANSQ-AQGR 914
DB 1174 GGARENGTLGAPNFAVGTGSIKVFYELIRDSNODYVLTRTNDVNVVTTAVGNSAIANAP 1233
QY 915 DLLQTL---LIDSHDAGYARKMID-----ATSANEITQOLNATATTTLNINIALEHKTSG 965
DB 1234 GVSQNTSRLESTNTAAYNNMLLAKOPSDVATEFVGAIATDTSAAVTTVN--LNDTOKTQD 1291
QY 966 LQTLSLSANMILNSRLNLSRRHTNHDGFAKRLQALQKQKFALESAAEVLVQFAPKYE 1025
DB 1292 L-----LSNRL---GTYRLYLSNAETSDVAGSATGAVSGDEAEVSYGVWAKPFNIAEQDK 1344
QY 1026 KPTNVWANAIGTSLNNGSNASLYGTSGVD---AYLNGQVEAIVG----- 1068
DB 1345 K-----GGIA---GYKAKTTGVVVGDLTTLASDNLMIGAAGITKTDIKHQDYKKGD 1392
QY 1069 --CGFGYGYSSFNRRNSLNSGANNTNFGVYSRIFANQHEFDEFAQALGSDOSSLNFKS 1126
DB 1393 KTDINGLSFLSGYSGQOLVKNFQAGNSIETLNVKVSQSQRFFESNGKMSQIAAGNYD- 1451
QY 1127 ALLQDLNOSVHYLAASAATRASGYDFAPFRNALVLKPSVSVNHLGTSNFKSNSTNOV 1186
DB 1452 -----NMTF-----GGNLIIFYDYNAMPNVLV-TPMAGLSYLLKSSNENYKGTGT-V 1496
QY 1187 ALKNGSS 1193
DB 1497 ANKRINS 1503

RESULT 11
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE: 16-MAR-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 4.4%; Score 292; DB 4; Length 1529;

Best Local Similarity 20.0%; Pred. No. 3e-11;
Matches 261; Conservative 161; Mismatches 482; Indels 402; Gaps 54;

QY 134 VDMK-----DAVGYTLGLRNFVG-----GO-----LDVNNQKATL 165
DB 360 VDAKEWLLDPDDVSLETLSGRNNTGENOGYTTGDGTEKSPKGNISKPTLTNSTLEQL 419
QY 166 RLQGF-----NGNSFTSY-----KDSADRTTRVDENAKNISIDNEVE 202
DB 420 RRGSVNITANNRIYVNSINLSGLTLTKRDGVKINGDITSNENGL-TIKAGSWD 478
QY 203 INRVGSGAG-----RKASSTVLTLOASEGITSOKNAEISLYDGT 243
QY 479 VHKNTLTGTGFLNIVAGDSVAFEREGRKARNATDAQITAGGT--ITVKKDDKQFRFNVS 536
QY 244 LNLASSSVKLMGN-----VWGRLOYVG-----AYLAPSYSTIN 277
DB 537 INGTGKGLKFIANNFTHKFGELNISIVTINOTTKDKVYWNASKDSYWNVSLSLTN 596
QY 278 TSKVTGEVNF-----NHLTVGDKNAQAAGIIAN-----KKTNIGTLDLWQSGAGLNIITAPPG 329
DB 597 TVQKFTFIKFDVDSGNSQDLRSRRSFAGVHENGIGKTN-----FNIGANAKA 645
QY 330 GYKDKPNTTPOSGAKNDKNSAKNDKQ-----SSQNSNTQVNPNSAQKTEV 380
DB 646 LFLKXPN-----AATDPKELPTIFNANITATGNSDSSVNFDIHANLTSRA 691
QY 381 OPTQVIDGPEAGGKDVVNINRINTNA-----DGTIRVGGFKASL--TTNAHLHIGKG 433
DB 692 AGINWDSNITGGLDFSTISHNRNNAEIKKDLTINATGNSFSLKQTKDSFYNEYSKHA 751
QY 434 VNLSNQASGRSLIVENLT---GNITVDGPLRVNNVGVYALAGSSANFEFFAGTDTKNGT 490
DB 752 INSSH-----NLTLGGNVTLGGE-NSSSITGNITNKNANTVLTQADTSNSN-T 799
QY 491 ATFNNDISLGRV-----NLKVDAAHTANFKGIDTGNNGFNLTDFSGVTDKVNINKLITAS-T 546

DB 800 GLKKRTLTLGNISVEGNLSLTGANANIV-----GNLSIAEDSTFKGEASDNLNITGTFT 853
QY 547 NVAVKFNELIVKNGISVGEYTHFSDIGSQSRINTVRLTGT-RSLFSGGVFKGG 605
DB 854 NNGTANINQGVKLOG-----DINKGGLNITTNASGTQKTIINGNITNEKG 902
QY 606 EKLVIDEYFYPWNYFDARNIKNVEITNKLAFGPOGSPWGTCKLMFNNTLGNVMDYS 665
DB 903 D-----LNKNIKADAEIQIG-----GNIS-----Q 923
QY 666 QFSNLTIOGDFIN--NOGTINYLVRGKVATLSVGNAAAMFNNIDISA-----TGFIKP 718
DB 924 KEGNLTISSDKVNITNQITIKAGVEGRSDSSEANANLTQIKELKLAGDLNISGFKA 983
QY 719 LKINSAQDL-INKTEHVLLKAKIIGYGV---STGTNGISVNVNLEQFKERLALYNNN 774
DB 984 EITAKNGSDLTIGNAGSGGNADAKKVTDFDKVDSKISTDG-HNVTLSSEVKTSGNSNAGN 1042
QY 775 RMDTCVVRNTDDIKACGMAIGDQSMVNNPDNYKYLIGKAWKNIGISKTAGNSKISVYILG 834
DB 1043 DNSGTGLTISAKDV-----TVNN-----NVTSHKTINISAAA---G 1074
QY 835 NSTPTENGNTNLPN---TTSNARSANNAALQAAPSAQPSATPNLVAINOHDFGTIE 890
DB 1075 NVTKE--CTTINATTGSVEVTAQNGTIKGNITSQNVTV---TATENLVTTENAVINATS 1129
QY 891 SVFELANRSKDI-----DLY 906
DB 1130 GTVNI STKTGDIKGGIESTSGNVNITASNTLKVSNITGODVTVTADAGALTTTAGSTIS 1189
QY 907 ANSG-----AQGRDLI-----OTLLIDSHDAGYARKMID 935
DB 1190 ATTGNANITTKTDINGKVESSGSVTLVATGATLAVGNISGNTVITA-DSKLTSTVG 1248
QY 936 AT-----SANEITKOLNTATTTLNNIASLEHKTGLQTLSTLSNAM 975
DB 1249 STINGTNSVTTSSQSDIEGTISGNTVNVTAFTGDLTIGNSAKVEAK-NGAATLTAESGK 1307
QY 976 ILNSRLVNLRRHTNHIIDFAKRLQALKDOKFASLESAAEVLQFAPKYEKPTNVANAI 1035
DB 1308 L-----TTQCGSSITSSNGOTTLTAKDSSIAGNINAANVTLTGTTTGDGSKINAT 1360
QY 1036 GGTSLNNGSNASLYGTSGVDAYLNGQVEATVGGFSGSYGSSFN-----NRANSLNSGAN 1090
DB 1361 SGTLTINAKDALDGAASGDRTVVNNATNASGSGNVTAKTSSSVNITGDLNTINGLNIIE 1420
QY 1091 NTNFGVYRIFANQHEFDE-AQAGLGDSSSLNFKSL--LQDLNOSYHYLAISAATRA 1147
DB 1421 NGRNTVRLR---GKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSD-----ERETLA 1470
QY 1148 SYGYDFAFFRNALVLKPSGVSYNHLGSTNFKSNSTNOVALKNGSS 1193
DB 1471 KLGVSAVR-----VEPNNAITVN--TONEFTTKPSSQVITISEGKA 1509

RESULT 13
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 4.3%; Score 287.5; DB 2; Length 1338;

Best Local Similarity 20.3%; Pred. No. 4.9e-11;
Matches 272; Conservative 165; Mismatches 539; Indels 365; Gaps 58;

QY 10 INRPLVSLALVGLVSTPQSHAAFFTVLIPAIVGIGIATGAVTVSGLLSLMGLKQAE 69
DB 186 IEAPSASRVELGA-----DRNSHSAEVIKVTLLKNNSTLTLNT-TISNL-----LKAH 235
QY 70 EAKTPDKPKVRIQAGKGFNEFPNKEYDLYRLLSSKIDGGWDGNAARHYYWKGQO 129
DB 236 VVNITARRKLT-----NSSISIERG-----SHLILHSEGGQGG 270
130 NKLVDKMDAVGTVL-----SGLRFNFTGGDLVNM-----K 162
271 VOIDKDTSEGNLTIYSGGVVDVHKHNTLGSGLNLTITTEGDIAPEDKSGRNNLTITAQ 330
QY 163 ATRLGQFNGNSFT-----SYKDS-ADRTTRVDFNAKN-----ISIDNFEVNN 205
DB 331 GTITSGNSGFRFNNVSLNSLGGKLSFTSDREDRGRRTKGNISKNKFDGTLNLSGTVDISM 390
QY 206 RVG---SCAGRKASST---VLTQASEGI-----TSDKNAEIS---LYDQAT 243
DB 391 KAPKVSFYRDKGRTYVNWTTLVNTSGKFNLSIDTSGSGSTGSPSIRNAELNGITFNKAT 450
QY 244 LNLASSSVKLMGNVWMLQVGVAYLAPSYSTINTSKVT-----GEVNFNHLTVGDK 295
DB 451 FNIAQGTANFS-----IKASIPFFKSNAYALFNEDISVGGSGSVNFK-LNASS 500
QY 296 NAAQAGIIANKTKTIGTLDLQWAGLNIAPPEGYKDKPNNTPSQSAKNDKRESAKND 355
DB 501 NIOTPGVII-KSQNFV-----SGSTLNLKAEQ-----STETAFSIENDLNLATGG 547
QY 356 KQESSQNNSTQVNPNSAQKTEVQTVIDGPFAGKQTV-----VNINRI 403
DB 548 NITRQVEGTDSDRVNKGVAANKNT-----FKGNTIFFGSKATTEIKGNVTINK- 597
QY 404 NTNADGTRVGGFKASLTNAAHHLHCKGGVNLNQASGRSLIVENTGNITVDGPLRVN 463
DB 598 NTNA--TLR-----CANFAENKSPUNIAGVINNCLNLTAGSII--NIAGNLTVSKGANLQ 649

QY 464 NOVGGYALAGSSANFEK-AGTDTKNGTATFNNDISLGRFVNKLKVDARTANFKGIDTNG 522
DB 650 -----ALTNYTFNVAGSPDNNGAS-----NISIARGGAKFKDINN--- 684
QY 523 GFNTLDFSGVTDKVINIKLITASTNAVAVKNFINELIVKTNIGISVGEYTHFSPEDIGSQR 582
DB 685 -----TSSLNITNTSDTYRTIIR-----GNISNKGSLNIDK 718
QY 583 INTVRLGTGRSLFSGGVKFKGEKLVIDFYYSPWNYFDARNIKNVEITNKLAF--GPO 640
DB 719 KSDAEIOIG-----GNISOKEGNLTISS-----KVNITNQIITIKAGVE 757
QY 641 GSPMGTSKLMFNLTLCQNAVMDYSQFSNLTIOGDFINNOCTINYLVRGKVKVATLSYVNA 700
DB 758 GGRSDSSEAEANLTI-----QTKELKLAGD-LNISGFNKAEITAKNGSDLTIGNA 807
QY 701 AA-----MMFNNDI-----DSATGFYKPLIKNSA 725
DB 808 SGNADAKKVTDFKVKDSKISTDGHNVTLNSEVKTSGSSNAGNDNSTGL-----TISA 861
QY 726 QDLIKN--TEHVLLKAKIIGYGNVST-----GTNGISNVNLEEQFKERLALYNNNR 775
DB 862 KDVTVNNVTSKHTINISAAA-GNVTTEKGTINATTGSEVETAQN-----GTIKGNIT 914
QY 776 MDTCVVRNTDDIKACGMAI-----GDQSMVNNPNKYLIKAKWKNIGIKSTANGSKISV 830
DB 915 SONVTVTATENLVTTENAVINATSGTVNISTPKTGDKGIGIESTSGNVNITASGNTLKVSN 974
QY 831 YVLGNSTPTENGNTNLP-----TNTTSNARSANNAQAOPAPQPSATPNLVAINOHDF 886
DB 975 ITGQDVTVTADAGALTTTAGSTISATTGANTITTKTGDKINGKVESSGSVTLVATG---- 1030
QY 887 GTIESVELANRSKIDITLVANSAGORDLLOTL-----LIDSHDAGYARKMIDAT-SAN 940
DB 1031 ----ATLAVGNISGNTVTTIADSGKLTSTVGSTINGNSVTSQSG-----DISGTISGN 1082
QY 941 EITKQLTATTTLNIAASLEHKTSGLOTLSLSNAMILNSRLVNSRRHTNIDHIDFAKRLQ 1000
DB 1083 TVNVYTAGDGLTIGNSAKVEAK-NGAATLTAESGKL-----TTQTGSSITSSNGOTTL 1134
QY 1001 ALKQDKFASLESAEVLVQFAPKEYEPTNVWANAIGTSLNNGSNASLYGTSAGVAYLN 1060
DB 1135 TAKDSSAGINAINAVNTLTGTLTTTGDSKINATSGTLTINAKDAKLDAAGSDRTVVN 1194
QY 1061 GOVEAIVGGEGSYGYSFN-----NRANSLNSGANNTNFGVYSRIFANQHEFDE-AQA 1114
DB 1195 ATNAGSGNVYAKTSSSVNITGDLNTINGLNIISENGRNTVRLR-----GREIDVYIOPG 1250
QY 1115 LGSQSSSLNFKSAL--LODLNQSYHYLAYSAATRASVGYDFAFPRNALVLKPSVGVSYNH 1172
DB 1251 VASVEEVEIAKRVLEKVKDLSDE-----BRETAKLGVSARE-----VEPNNAITVN- 1298
QY 1173 LGSTNFKSNSTNQVALKNGSS 1193
DB 1299 -TQNEFTTKPSSQVITISEGA 1318

RESULT 14

US-08-719-641-9
; Sequence 9, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Batenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-9

Query Match 4.3%; Score 287.5; DB 4; Length 1338;

Best Local Similarity 20.3%; Pred. No. 4.9e-11;

Matches 272; Conservative 165; Mismatches 539; Indels 365; Gaps 58;

QY 10 INRPLVSLALVCAVSIITQOASHAAFFTVIPIAVGGTATGAVTVSGLLSWGLKQAE 69
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
DB 186 IEAPSARVELGA-----DRNSHSAEVIKVLKKNNTSLTTLTNT-TISNL-----LKSAH 235
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 70 EANKTPDKPKVYRIQAGKGFNEPKYDLYRSLLESLKIDGWDGNAARHYVWKGGO 129
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
DB 236 VVNIARRLTV-----NSSISIERG-----SHLIHSEGGGGQ 270
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 130 NKLEVDKMDAVETYL-----SGLRFTGGDLVNMQ-----K 162
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
271 VQIDKDTSEGGLTIYSGWVDVHKNTILGSGFLNITTKEGDIAFDKSGRNNLTITAQ 330
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 163 ATLRLQFQNGNSFT-----SYKDS-ADRTTRVDFNAKN-----ISIDNFVEINN 205
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
DB 331 GTITSGNSGFRFNNVNSLGLGKLSFTDSREDRGRRTKGNISNKFDTGLTINISGTVDISM 390
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 206 RVG--SGAGRKASST---VLTLOASEGI-----TSDKNAEIS--LYDGAT 243
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
DB 391 KAPKYSWFYRDKGRTVWNTTLNVTSGKENLSIDTSGSGTSPSIRNAELNGITFNKAT 450
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 244 LNLASSVKLMGVNMGRIQYVGAYIAPYSYTIINTSKVT-----GEVNFNHLTVGDK 295
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
DB 451 FNIAGOSTANFS-----IKASIMPKSNANYALFNEIDISVSGGGSVNFK-LNASSS 500
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 296 NAAQAGLIANKTNGTLDLWOSAGLNIAPPEGYKDKPNWTPSOSGAKNDKNEAKND 355
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
DB 501 NIQTGCVIIT-KSQNFNV-----SGGSTLNLKABG-----STETAFSIENDLNLATGG 547
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 356 KQESSNNNSNTQVNPNSAOKTEVQPTQVIDPAGGKDTV-----VNIINRI 403
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
DB 548 NITIROVEGDSRVNKGVAANKNIT-----FKGNNITFGSQKATTEIKGNVTINK- 597
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 404 NTNADGRTIRVGGFKASLTNNAHLHIGKGVNLSNOASGRSLIVENLTGNITVDGPLRVN 463
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :

RESULT 15

US-08-617-697-9

; Sequence 9, Application US/08617697

; Patent No. 5977336

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

Db 598 NTNA--TLR-----GANFAENKSPLNAGNVINNGNLTAGSII--NIAGNLTVSKGANLQ 649
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 464 NOVGGYALAGSSANFEFK-AGTDTKNGTATFNNDISLGRFVNKLKVDHAHTANFKGIDTNG 522
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 650 -----AITNYTFNVAGSFDNNGAS-----NISIARGGAKFKDINN--- 684
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 523 GFNTLDFSGVTDKVNINKLITASTNVAVKNFNIENELIVKTNGISVGYTHFSEDIGSQSR 582
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 685 -----TSSLNITTSNDSITYRIIK-----GNISNKSGLDLNIDK 718
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 583 INTVRLTGTSLFSGGVKPKGGEKLVDFEYYPSPWNYFDMARKIKNVEITNKLAFF--GPQ 640
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 719 KSDAEIQIG-----GNISQKEGNLTSSD-----KVNITNQITIKAGVE 757
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 641 GSPWGTSKLMFNLTGONAVMDYSQFSNLTIOGDFINNOGTINYLVRGKVKVATLSVGN 700
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 758 GGRSDSSEANALTI-----QTKELLAGD-LNISGFNKAETAKNGSDLTIGNA 807
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 701 AA-----MMFNNDI-----DSATGYFKPLIKINSA 725
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 808 SSGNADAKKVTDPKVKDSKISTDGHNVTLNSEVKTSGSSNAGNDNSTGL-----TISA 861
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 726 QDLIKN---TEHVLLKAKIIGYGVST-----GTNGISNVNLEEOFKERLALYNNNR 775
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 862 KDVTVNNVTSHKTIINISAAA-GNVTTKEGTTINATTGSEVEVTAQN-----GTIKGNIT 914
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 776 MDCVVRNTDDIKACGMAL-----GQOSMVNPDNPKYLIGKAKKNIGISKTANGSKISV 830
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 915 SQNVTVTATENLVTTENAVINATSGTVNISTKTDGDKGGIESTSGNVNITASNTLKVSN 974
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 831 YYLGNSTPTENGNTNLP-----TNTTSNARSANALAOAQAQPSATPNLVAINGHDF 886
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 975 ITGQDVTVTADAGALTTTAGTISATIGNANIITKTGDINGKVESSESGSVTLVATG---- 1030
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 887 GTIESVFEIARNSKIDITLYANSAGQGRDLLQTL-----LIDSHDAGYARKMIDAT-SAN 940
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1031 -----ATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVTTSSQSG-----DIEGTISGN 1082
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 941 EITKQLNTATTTLNNTASLEHKTSGLOTLSNAMILNSLRHTNHIDSPARLQ 1000
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1083 TVNVTASTGLDTIGNSAKVEAK-NGAATLTAESGKL-----TTQTGSSITSSNGQTTL 1134
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 1001 ALKQDFASLESAAEVLYQAPKYEKPNTVWANAIGGTSLNNGSNASLYGTSAGVDAYLN 1060
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1135 TAKDSSIAGNINAAVTLNTTGLTTTGDSKINATSGTLINAKDAKLDCAASDRTVVN 1194
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 1061 GQVEAIVGGFGSYGYGSFN-----NRANLSNGANNFTFGVYSRIFANQHEFDPE-AQGA 1114
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1195 ATNAGSGNVTAKTSSSVNITGDLNTINGLNIISENGRNTVRLR-----GKEIDVKYIQPG 1250
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 1115 LGSQDSLSNFKSAL--LQDLNQSHYLAASAATRASVGYDFAFFRNALVLKPSVGVSYNH 1172
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1251 VASVEEVIEAKRVLEKVKDLSD-----ERETLAKLGVSAVRF-----VEPNNAITVN- 1298
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
QY 1173 LGSTNFKSNSTNOVALKNSS 1193
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1299 -TONEFTTKPSSQVTTISEGKA 1318
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :

RESULT 15

US-08-617-697-9

; Sequence 9, Application US/08617697

; Patent No. 5977336

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:27:24 ; Search time 29.02 Seconds
(without alignments)
4291.240 Million cell updates/sec

Title: US-09-921-157-3
Perfect score: 6671
Sequence: 1 MEIQOTHRKNRPLVSLALV.....HNLISNIGHFASNLGMRYSF 1296

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6303.5	94.5	1287	2 B53739	vacuolating cytoto
2	6222.5	93.3	1291	2 S44983	vacuolating cytoto
3	6220	93.2	1290	2 G4630	vacuolating cytoto
4	6108	91.6	1288	2 E71884	vacuolating cytoto
5	1695.5	25.4	513	2 A53739	hypothetical vacuo
6	674	10.1	2902	2 C71953	toxin-like outer m
7	648.5	9.7	2893	2 A64556	toxin-like outer m
8	642.5	9.6	2529	2 B64635	toxin-like outer m
9	623	9.3	2399	2 H71879	toxin-like outer m
10	583	8.7	1943	2 B64596	toxin-like outer m
11	580.5	8.7	3194	2 D71917	toxin-like outer m
12	335	5.0	1655	2 E97835	hypothetical prote
13	314.5	4.7	1300	2 S07575	outer membrane pro
14	313	4.7	2249	2 A41477	190K surface anti
15	312.5	4.7	2020	2 C48399	ABC-type transport
16	310.5	4.7	2021	2 A97859	190-KDa cell surfa
17	310	4.6	1286	2 S28634	adhesin AIDA-I pre
18	310	4.6	1651	2 JC1340	outer membrane pro
19	306	4.6	1643	2 D71630	outer membrane pro
20	303.5	4.5	2340	2 B71704	cell surface anti
21	297.5	4.5	1270	2 E85649	hypothetical prote
22	297.5	4.5	3705	2 A00123	probable autotrans
23	296.5	4.4	1268	2 B99789	hemagglutinin/hemo
24	291	4.4	1268	2 A80204	conserved hypotet
25	291	4.4	1335	2 T17508	glycoprotein Vp260
26	291	4.4	1645	2 JN0896	crystalline surfac
27	290	4.3	1519	2 S41525	major ring-forming
28	283	4.2	1541	2 A37023	IgA-specific metal
29	282	4.2	1327	2 B90674	AIDA-I adhesin-lik

ALIGNMENTS

RESULT 1
B53739

vacuolating cytotoxin precursor - Helicobacter pylori (strain ATCC 49503)
C:Species: Helicobacter pylori
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: B53739; A38137
R:Cover, T.L.; Tummuru, M.K.R.; Cao, P.; Thompson, S.A.; Blaser, M.J.
J. Biol. Chem. 269, 10566-10573, 1994
A:Title: Divergence of genetic sequences for the vacuolating cytotoxin among Helicoba
A:Reference number: A53739; MUID:94193753
A:Accession: B53739
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1287 <CO>
A:CROSS-references: GB:U05676; MID:9471727; PIDN:AAAL17657.1; PID:9471729
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Cover, T.L.; Blaser, M.J.
J. Biol. Chem. 267, 10570-10575, 1992
A:Title: Purification and characterization of the vacuolating toxin from Helicobacter
A:Reference number: A38137; MUID:92268100
A:Accession: A38137
A>Status: preliminary
A:Molecule type: protein
A:Residues: 34-56 <CO>
A:Note: sequence extracted from NCBI backbone (NCBIP:103729)
C:Genetics:
A:Gene: vacA
C:Keywords: cytotoxin
F.1-33/Domain: signal sequence #status predicted <SIG>

Query Match 94.5%; Score 6303.5; DB 2; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;

QY	1	MEIQOTHRKNRPLVSLALVSLTPQOSHAAFTTVPVIAVGGIATGTAVGVSG	60
DB	1	MEIQOTHRKNRPLVSLALVSLTPQOSHAAFTTVPVIAVGGIATGTAVGVSG	60
QY	61	LSWGLKQAEANKTPDKPKVWRIQAGKGFNFPPNKKEYDLYRSLSSKIDGGWDMGNAAR	120
DB	61	LSWGLKQAEANKTPDKPKVWRIQAGKGFNFPPNKKEYDLYRSLSSKIDGGWDMGNAAT	120
QY	121	HYWVGKQONKLEVDKMDKAVGTYTSLGLNFTGGDLVNNQKATLRGQFNQNSFTSYKD	180
DB	121	HYWVGKQONKLEVDKMDKAVGTYTSLGLNFTGGDLVNNQKATLRGQFNQNSFTSYKD	180
QY	181	SADRTTRVDFNAKINSIDNFVEINNRVSGAGRKASSTVLTQASGITSDKNAEISLYD	240
DB	181	SADRTTRVDFNAKINSIDNFVEINNRVSGAGRKASSTVLTQASGITSSKNAEISLYD	240
QY	241	GATNLASSSVKLMGNVMGRQLQYVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA	300

probable beta-barr
ydeK protein - Esc
hypothetical prote
hypothetical prote
hypothetical prote
IgA-specific metal
hypothetical prote
hypothetical prote
extracellular seri
filamentous hemagg
hypothetical prote
hypothetical prote
high-molecular-wei
cell surface anti
surface protein Xf
probable adhesin h

241	Db	:	GATNLANSVYKLVGNVWGRLOYVAYLAYSSTINTSKVTGEVFNHLTVGDHNAQA	300
301	Qy		GIANKTKWIGTLDLWQSAGLNIIAPEGGYKDKPNNTPSQSGAKNDKNSAKNDKQESS	360
301	Db		GIASNKTHIGTLDLWQSAGLNIIAPEGGYKDKPNNTPSQSG-----AKNDKQESS	352
361	Qy		QNNSTQVNLPPNSAQKTEVQPTVIDGPFAGGDKTVVNIINRINTNADGRTIRVCGFKASL	420
353	Db		QNNSTQVNLPPNSTQKTEVQPTVIDGPFAGGDKTVVNIIDRINTKADGTIKVGGFKASL	412
421	Qy		TTNAAHLRIKGGVNLNQASGRSLIVENLTGNITVDGPLRVNNQVGYALAGSSANFEF	480
413	Db		TTNAAHLNIGKGVNLNQASGRTLLENLTGNITVDGPLRVNNQVGYALAGSSANFEF	472
481	Qy		KAGDTKNGTATFNNDISLGRFVNLKYDAHTANFKGIDTNGGENTLDFSGVTDKYNINK	540
473	Db		KAGVDTKNGTATFNNDISLGRFVNLKYDAHTANFKGIDTNGGENTLDFSGVTNKYNINK	532
541	Qy		LITASTNVAVKFNELIVKTNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV	600
533	Db		LITASTNVAVKFNELIVKTNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV	592
601	Qy		KPKGGEKLVIDFYYSPWNVEDARNIKVETNKLAFGQPGSPNGTSKLMFNLLTQGNA	660
593	Db		KPKSGEKLVIDFYYSPWNVEDARNIKVETTRKFASPTPENPGWTSKLMFNLLTQGNA	652
661	Qy		VMDYSQFSLTTQGGFINNOGTINYLVRGGKVATLSVGNAAAMFNNDISATGFYKPLI	720
653	Db		VMDYSQFSLTTQGGFINNOGTINYLVRGGKVATLVGNAAAMFNNDISATGFYKPLI	712
721	Qy		KINSAODLKNTHEVLLKAKIIGYGVSTGTNGISVNLSEQFKERLALYNNNRMDTCV	780
713	Db		KINSAODLKNTHEVLLKAKIIGYGVSTGTNGISVNLSEQFKERLALYNNNRMDTCV	772
781	Qy		VRNTDDIKACGMAIGDSWVNPNDNYKYLIKAKWNKIGISKTANGSKISVYI LGNSTPTE	840
773	Db		VRNTDDIKACGMAIGDSWVNPNDNYKYLIKAKWNKIGISKTANGSKISVYI LGNSTPTE	832
841	Qy		NGGNTNLPTNTSNARSANNALQAQNAFPAQSPATPNLVAI NQHDFTGIESVPELANRSK	900
833	Db		NGGNTNLPTNTNARFASVALIKNAPFAH-SATPNLVAI NQHDFTGIESVPELANRSK	891
901	Qy		DIDTLYANSGAOGRDLQTLTLLIDSHDAGYARKMIDAYSANEITHKOLTATTTLNNIASLE	960
892	Db		DIDTLYANSGAOGRDLQTLTLLIDSHDAGYARTMIDAYSANEITHKOLTATTTLNNIASLE	951
961	Qy		HKTSGIQLTSLSNAMILNSRLVNSRRRTHNIDSFARKLQALKDQKFALESAAEVLYQF	1020
952	Db		HKTSSIQLTSLSNAMILNSRLVNSRRRTHNIDSFARKLQALKDQKFALESAAEVLYQF	1011
1021	Qy		APKYEKPTNVWANAIGTGTSLNNGSNASLYGTSAGVDAYLNGQVFAIVGGFGSGYSSFNN	1080
1012	Db		APKYEKPTNVWANAIGGASLNGGNASLYGTSAGVDAYLNGQVFAIVGGFGSGYSSFNN	1071
1081	Qy		RANSLNSGANNTFNGVYSRIFANQHEFDEAQAQGLSDOSSLNFKSALLQDLNQSYYHLA	1140
1072	Db		QANSLNSGANNTFNGVYSRIFANQHEFDEAQAQGLSDOSSLNFKSALLRDLNQSYYHLA	1131
1141	Qy		YSAATRASYGYDFAFRNALVLKPSGVSYVNHLSGTNFKSNSTNQVALKNGSSQHFLFNA	1200
1132	Db		YSAATRASYGYDFAFRNALVLKPSGVSYVNHLSGTNFKSNSTNKVALSNGSSQHFLFNA	1191
1201	Qy		SANVEARYYIGDTSFYWMNAGVLOEFAHFVHSGVSNNAASLNTPKVNAARNPLNTHARVMGGE	1260
1192	Db		SANVEARYYIGDTSFYWMNAGVLOEFAHFVHSGVSNNAASLNTPKVNAARNPLNTHARVMGGE	1251
1261	Qy		LKLAKEVFLNLGVYDLHNLISNIGHFASNLGMYSF	1296
1252	Db		LKLAKEVFLNLGVYDLHNLISNIGHFASNLGMYSF	1287

RESULT 2

S44983
 vacuolating cytotoxin precursor - *Helicobacter pylori* (isolate 185-44)
 C:Species: *Helicobacter pylori*
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
 C:Accession: S44983; S44102
 R:Schmitt, W.; Haas, R.
 Mol. Microbiol. 12, 307-319, 1994
 A:Title: Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural
 A:Reference number: S44983; MUID:94335650
 A:Accession: S44983
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1291 <SCH>
 A:Cross-references: EMBL:Z26883; NID:g472941; PIDN:CAA81528.1; PID:g472942
 C:Keywords: cytotoxin
 F:1-33/Domain: signal sequence #status predicted <SIG>
 F:34-1291/Product: vacuolating cytotoxin #status predicted <MAT>

Query Match	93.3%	Score 6222.5	DB 2	Length 1291
Best Local Similarity	92.7%	Pred. No. 0		
Matches 1202	Conservative 51	Mismatches 38	Indels 5	Gaps 1
QY	1	MEIQOTHRKNRPLVSLALYCALVSIPTQOSHAAFFTVIIPAIVGGIATGTAVGTYSGL	60	
DB	1	MEIQOTHRKNRPLVSLALYCALVSIPTQOSHAAFFTVIIPAIVGGIATCAAVGTYSGL	60	
QY	61	LSWGLKQAEANKTPDKPDKVRIQACKGFNEFPNKEYDLYRSLSSKIDGGWDGNAAR	120	
DB	61	LWGLKQAEANKTPDKPDKVRIQACKGFNEFPNKEYDLYKSLSSKIDGGWDGNAAR	120	
QY	121	HYWYKGGQONKLEVDKMDVAGTVTSLGLRNFTHGGLDVNMOKATLRLQFNGNSFTSYKD	180	
DB	121	HYWYKGGQONKLEVDQMNAVGTYNLSGLINFTGGDLVNMOKATLRLQFNGNSFTSYKD	180	
QY	181	SADRTTRVDFNAKNISIDNFVEINNRVSGSAGRKASSTVLTQASEGITSDDKNAEISLYD	240	
DB	181	SADRTTRVDFNAKNILIDNFLEINNRVSGSAGRKASSTVLTQASEGITSRENAEISLYD	240	
QY	241	GATILNLAASSVKLMGNVWMCRLQVVGAYLAPSYSTINTSKVTGEVNFHLLTVGDKNAQA	300	
DB	241	GATILNLAASSVKLMGNVWMCRLQVVGAYLAPSYSTINTSKVTGEVNFHLLTVGDHNAQA	300	
QY	301	GIATKKTNTIGTLDLWQASGLNIITAPPEGGYKDKPNNTPSQSGAKNDKESAKNDKQESS	360	
DB	301	GIATKKTNTIGTLDLWQASGLNIITAPPEGGYKDKPNNDKPS- - - -NTQNNAKNDKQESS	355	
QY	361	QNSNSTQVIINPPNSAQTEVQPTQVIDPPAGGKDTVVNINRINTNADGTRVGGFKASL	420	
DB	356	QNSNSTQVIINPPNSAQTEIQPTQVIDPPAGGKNTVYNINRINTNADGTRVGGFKASL	415	
QY	421	TTNAAHLHIKGGVNLNSQASGSLIVENITGNITVDGLRVNNOVGGYALAGSSANFEE	480	
DB	416	TTNAAHLHIKGGVNLNSQASGSLIVENITGNITVDGLRVNNOVGGYALAGSSANFEE	475	
QY	481	KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFGNTLDFSGVTDKVNINK	540	
DB	476	KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFGNTLDFSGVTKVNINK	535	
QY	541	LITASTNVAVKNFNEINELIVKTINGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV	600	
DB	536	LITASTNVAVKNFNEINELVKTNGVSVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV	595	
QY	601	KFKGGEKLVIDEFYSPWNTFDARNIKNVEITNKLAFPGQSGPWGTSKLMFNLLTLQONA	660	
DB	596	KFKGGEKLVINDEFYAPWNTFDARNIKNVEITNKLAFPGQSGPWGTAKLAFNLLTLQONA	655	
QY	661	VMDYSQFSNLTIGDFTNNOGTNYLVRGCKVATLSVGNAAAMFNNDIDSATGYFKPLI	720	
DB	656	VMDYSQFSNLTIGDFTVNNOGTINLYRGQVATLVNGNAAAMFFSNVDSATGYFKOPLM	715	
QY	721	KINSAODLIKNTHEVLKAKIIGYGVNSTGTNGISNVLBEQFKERIALYNNNNRMDTCV	780	

Db 716 KINSAODLTKNKEHVLLKAKIIGYGNVSAGTDSIANVNLIEQFKERLALYNNNRMDICV 775
QY 781 VRNTDDIKACGMAIGDQSMVNNPDNYKYILGKAWKNIGISKTAGSKISVYILGNSTPTE 840
Db 776 VRNTDDIKACGTAIGNQSMVNNPENYKYILGKAWKNIGISKTAGSKISVHYILGNSTPTE 835
QY 841 NGGNTTNLPTNTTSNARSANNAALQAAPQASATPNLVAIQHDFGTIESVFELANRSK 900
Db 836 NGGNTTNLPTNTTKYRFASALIKNAPFARYSATPNLVAIQHDFGTIESVFELANRSN 895
QY 901 DIDTLYANSAGORDLLOLLDSDHDAGYARKMIDATSANEITKQLTATTTLLNNIASLE 960
Db 896 DIDTLYANSAGORDLLOLLDSDHDAGYARKMIDATSANEITKQLTATTTLLNNIASLE 955
QY 961 HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLQYF 1020
Db 956 HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLQYF 1015
QY 1021 APKYEKPTNWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1080
Db 1016 APKYEKPTNWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1075
1081 RANSLNSGANNTNFGYYSRIFANQHEFDEAOGALGSDQSSLNFKSALLQDLNQSYHYLA 1140
Db 1076 QANSLNSGANNTNFGYYSRIFANQHEFDEAOGALGSDQSSLNFKSALLQDLNQSYHYLA 1135
QY 1141 YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGSSQHLFNA 1200
Db 1136 YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGASSQHLFNA 1195
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFQFAHVGSNNAASLNTFKYNAARNPLNTHARVMMGE 1260
Db 1196 SANVEARYYYGDTSYFYMNAGVLQEFQFAHVGSNNAASLNTFKYNAARNPLNTHARVMMGE 1255
QY 1261 LKLAKEVFLNLGVVYLNHLNLSNIGHFASNLGMRYSF 1296
Db 1256 LKLAKEVFLNLGVVYLNHLNLSNIGHFASNLGMRYSF 1291

RESULT 3
G64630
vacuolating cytotoxin precursor - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: G64630
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
Reference number: A64520; MUID:97394467
A:Accession: G64630
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1290 <TOM>
F.1-33/Product: vacuolating cytotoxin #status predicted <SIG>
F.34-1290/Product: vacuolating cytotoxin #status predicted <MAT>

Query Match 93.2%; Score 6220; DB 2; Length 1290;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1207; Conservative 38; Mismatches 45; Indels 6; Gaps 2;

QY 1 MEIQOTHRKNINPLVSLALVGALVSTTPQOASHAAFTTVIIIPAVGGIATGATGAVTSG 60
Db 1 MEIQOTHRKNINPLVSLALVGALVSTTPQOASHAAFTTVIIIPAVGGIATGAVTSG 60
QY 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNFPKNKYDLYRSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNFPKNKYDLYRSLSSKIDGGWDMGNAAT 120

QY 121 HYWVGQGNQKLEVDKMDKDAVGTYYTLGLRNFRTGGDLVDNNMQKATRLRGQFNQNSFTSYKD 180
Db 121 HYWVGQGNQKLEVDKMDKDAVGTYYTLGLRNFRTGGDLVDNNMQKATRLRGQFNQNSFTSYKD 180
QY 181 SADRTTRVDFENAKNISIDNFEVNNRVGSGAGRKASSTVLTLOASBEGITSDKNAEISLYD 240
Db 181 SADRTTRVDFENAKNIILDNFLEINNRVGSAGRKASSTVLTLOASBEGITSSKNAEISLYD 240
QY 241 GATLNLASSSVKLMGNVWGRLOQYVGAYLAPSYSTNTSKVTGEVNFHLLTVGDKNAAQA 300
Db 241 GATLNLASSSVKLMGNVWGRLOQYVGAYLAPSYSTNTSKVTGEVNFHLLTVGDKNAAQA 300
QY 301 GIITANKKTIGTLDLWQASAGLNIIPPEGYKDKPNNTPSQSGAKNDKNESAKNDQESS 360
Db 301 GIITANKKTIGTLDLWQASAGLNIIPPEGYKDKPKDPS-----NTQNNANNQNSA 355
QY 361 QNNSNTQVINPPNSAOKTEVQPTQVIDGPFAGCKDVIWVNNRINTNADGTIRVGGPKASL 420
Db 356 QNNSNTQVINPPNSAOKTEQPTQVIDGPFAGCKDVIWVNNRINTNADGTIRVGGPKASL 415
QY 421 TTNAHLHIGKGVNLSNQSASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEP 480
Db 416 TTNAHLHIGKGVNLSNQSASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEP 475
QY 481 KAGTDKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNNGFNTLDFSGVTDKVNINK 540
Db 476 KAGTDKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNNGFNTLDFSGVTDKVNINK 535
QY 541 LITASTNVAVKPNINELIVKNGISVGEYTHFSEIDIGSOSRINTVLETGTSLPSGGV 600
Db 536 LITASTNVAVKPNINELIVKNGISVGEYTHFSEIDIGSOSRINTVLETGTSLPSGGV 595
QY 601 KFKGGEKLVITDEPYSPWNYFADARNIKNVEITNKLAFPGPOGSPWGSKLMFNLLTGONA 660
Db 596 KFKGGEKLVITDEPYSPWNYFADARNIKNVEITNKLAFPGPOGSPWGSKLMFNLLTGONA 655
QY 661 VMDYSQFSNLTIQGDFINNOGTINYLVRGKVATLSVGNAAAAMFNNDISATGPKPLI 720
Db 656 VMDYSQFSNLTIQGDFINNOGTINYLVRGKVATLSVGNAAAAMFNNDISATGPKPLI 715
QY 721 KINSAODLTKNKEHVLLKAKIIGYGNVSAGTDSIANVNLIEQFKERLALYNNNRMDTCV 780
Db 716 KINSAODLTKNKEHVLLKAKIIGYGNVSAGTDSIANVNLIEQFKERLALYNNNRMDICV 775
QY 781 VRNTDDIKACGMAIGDQSMVNNPDNYKYILGKAWKNIGISKTAGSKISVYILGNSTPTE 840
Db 776 VRNTDDIKACGTAIGNQSMVNNPDNYKYILGKAWKNIGISKTAGSKISVYILGNSTPTE 835
QY 841 NGGNTTNLPTNTTSNARSANNAALQAAPQASATPNLVAIQHDFGTIESVFELANRSK 900
Db 836 NGGNTTNLPTNTTSNARSANNAALQAAPQASATPNLVAIQHDFGTIESVFELANRSK 895
QY 901 DIDTLYANSAGORDLLOLLDSDHDAGYARKMIDATSANEITKQLTATTTLLNNIASLE 960
Db 896 DIDTLYANSAGORDLLOLLDSDHDAGYARKMIDATSANEITKQLTATTTLLNNIASLE 955
QY 961 HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLQYF 1020
Db 956 HKTSGLOTLSLNSAMILNSRLNLSRRHTNHIDSFARLQALKDQKFALESAAEVLQYF 1015
QY 1021 APKYEKPTNWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1080
Db 1016 APKYEKPTNWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFNN 1075
QY 1081 RANSLNSGANNTNFGYYSRIFANQHEFDEAOGALGSDQSSLNFKSALLQDLNQSYHYLA 1140
Db 1076 QANSLNSGANNTNFGYYSRIFANQHEFDEAOGALGSDQSSLNFKSALLQDLNQSYHYLA 1135
QY 1141 YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGSSQHLFNA 1200
Db 1136 YSAATRASYGYDFAPFRNALVLKPSGVSYNHLGTSNFKSNSTNQVALKNGSSQHLFNA 1194
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFQFAHVGSNNAASLNTFKYNAARNPLNTHARVMMGE 1260

Qy	601	KFKGGEKLVIDEEYYSPWNYFDARNITKNVEITNKLAFGQGSWGTSKLMPNNLTILGQNA	660
Db	594	KFKSGERKLVINDIEYYSPWNYFDARNYKNVEITRKFASSTPENPWGTSKLMPNNLTILGQNA	653
Qy	661	VMDYSQPSNLTIQDFTINNGTINYLVRGGKVATLVSGVNAAMFNNDIDSATGFYKPLI	720
Db	654	VMDYSQPSNLTIQDFTINNGTINYLVRGGKVATLVSGVNAAMFNNDIDSATGFYKPLI	713
Qy	721	KINSAQDLIKNTEHVLKAKIIGYGVNSTGTNGISNVNLEEOFKERLALYNNNNRMDTCV	780
Db	714	KINSAQDLIKNTEHVLKAKIIGYGVNSTGTNGISNVNLEEOFKERLALYNNNNRMDTCV	773
Qy	781	VRNTDDIKACGMAIGDOSMVNNPDNYKYLIGKAWKNIGISKTAGSKISVYYLGNSTPTE	840
Db	774	VRNTDDIKACGMAIGDOSMVNNPDNYKYLIGKAWKNIGISKTAGSKISVYYLGNSTPTE	833
Qy	841	NGGNTTLPNTTNSARSANALQAQNPAPQSPATPNLVAINQHDGTTESVFELANRSK	900
Db	834	NGGNTTLPNTTNNAHSAHYALVKNAPFAH-SATPNLVAINQHDGTTESVFELANRSK	892
Qy	901	DIDTFLYANSAQGRDLLQTLIDSHDAGYARKMIDATSANEITKOLNTATTTLNNTIASLE	960
Db	893	DIDTFLYTHSQAQGRDLLQTLIDSHDAGYARKMIDNTSTGEITKQNAATDALNNVASLE	952
Qy	961	HKTSGLOTLISLNMATILNSRLVLSRRRTHNIDISFAKRLQALKDOKFASLEAAEVLVYOF	1021
Db	953	HKQSGLOTLISLNMATILNSRLVLSRKHTNHINSFAORLQALKGQEFASLEAAEVLVYOF	1011
Qy	1021	APRYEKPNTVWANAIGTSLNNGSNASLVGTSGAGVDAYLVNGQVEATVGGFGSYGSSFPNN	1081
Db	1013	APRYEKPNTVWANAIGASLNSGNSASLVGTSGAGVDAFLNGVNEATVGGFGSYGSSFSN	1071
Qy	1081	RANSLNSGANNTPGVVYSRTFANQHEFDEAQCALGSDOSSLNFKSALLQDLNQSIVHYLA	1141
Db	1073	QANSLNSGANNANPVGYSRFFANQHEFDEAQCALGSDOSSLNFKSLTLLQDLNQSIVHYLA	1131
Qy	1141	YSAATRASVGYDFAFFRNALVLPKPSVGVSYNHLGSTNFKNSNTNQVALKNGSSQHLFNA	1201
Db	1133	YSATRASVGYDFAFFRNALVLPKPSVGVSYNHLGSTNFKNSNSQSOVALKNGASSQHLFNA	1191
Qy	1201	SANVEARYYYGDTSYFYMNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVVMGGE	1261
Db	1193	NANVEARYYYGDTSYFYHLAGVLQEFAGHFGSDNVASLNTFKINAARSPLSTYARAMMGGE	1251
Qy	1261	LKLAKEFELNLGVVYVYHLNLSNIGHFASNLGMRYSF	1296
Db	1253	LQLAKEFELNLGVVYVYHLNLSNASHFASNLGMRYSF	1288

RESULT 5

A53739

hypothetical vacuolating cytotoxin - Helicobacter pylori (strain ATCC 53729)

C:Species: Helicobacter pylori

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999

C:Accession: A53739

R:Cover, T.L.; Tummuru, M.K.R.; Cao, P.; Thompson, S.A.; Blaser, M.J.

J. Biol. Chem. 269, 10566-10573, 1994

A:Title: Divergence of genetic sequences for the vacuolating cytotoxin among Helicobacter strains

A:Reference number: A53739; MUID:94193753

A:Accession: A53739

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-513 <COV>

A:Cross-references: GB:005677; NID:9471730; PIDN:AAAL7658.1; PID:9471731

A:Note: This strain, designated 87-203, ATCC 53726 tox-, does not possess cytotoxicity

C:Genetics:

A:Gene: vacA

Query Match	25.4%	Score 1695.5;	DB 2;	Length 513;
Best Local Similarity	63.3%;	Pred. No. 9.3e-80;		
Matches 343;	Conservative	52;	Mismatches 84;	Indels 63;
				Gaps

Db 2629 KHPNMLIQGVGGASFSGGNGTLGLNVGVDRLVK---SVILGGYVAYGSGFN--GNI 2693
Qy 1085 LNSGANNTEGCVYGRIFANQHEFEAOGALGSDQSILNFKSALLQDLNOSYHVLAYSAA 1144
Db 2684 MHSLANVVDVGMVYARAFKRNFTLSANEYGGNASHINSNLSLLSVLNQRYNTWTTS 2743
Qy 1145 TRASYGDFAPFRNALVLPKPSVGSYNHLGNTNPKSNSTN---OVALKSSSSQHILFNA 1200
Db 2744 VNGNYGDFMFKOKSVVLKPOVGLSYHFGLSGMKGMQNPAYQQFVMSNPSNESVLT 2803
Qy 1201 SANVEARYYYGDTSYFYNNAGVQLQEF-----AHVGSNNAASLNTFKVNAARPLN 1250
Db 2804 NMGLESRYKFGKNSYYFVTLARGLDLILAKAGDNVVRFGENTL-----LYRKGIFN 2856
Qy 1251 THAVMMGGELKAKEVFLNGLVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1296
Db 2857 TFASVINGGEMHLWRLMYVNAVGLKMGLOQYQDINTGNGMRVAF 2902
RESULT 7
A:Accession: A64556
toxin-like outer membrane protein HP0289 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: A64556
R:Tomb, J.F.; White, O.; Kervage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64556; MUID:9739467
A:Accession: A64556
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2893 <TOM>
A:Cross-references: GB:AE000547; GB:AE000511; MID:g2313377; PIDN:AA007355.1; PID:g231338

Query Match 9.7%; Score 648.5; DB 2; Length 2893;
Best Local Similarity 23.5%; Pred. No. 7.1e-25;
Matches 318; Conservative 189; Mismatches 493; Indels 353; Gaps 57;

Qy 114 DMGNAARHYWYKGGQKLEVDKMDAVGTYSGLRNFTEGGDLVDNMQKATRLGQFNGN 173
Db 1724 DWKN-----INELIDNANNVQNFNGNTLIV-----GATQIGQDTN 1761
174 SFT-----SYKDSADRTTRV-----DFNAKNISIDNFEI 203
Db 1762 SAVVGGGLGYQTPCDYTDIVCQKFRGTYLGLLESSADLGYIDTTFNAKEIYLTGTLS 1821
Qy 204 NNRVSGGAGRKA---SSTVLTQLAQSEGITSDKNAEISLYDGALENLASSSVKL----- 253
Db 1822 GNWGTGGSASVTNSQTSLLNQCANIVSSQTDGIFSMGLGEGINKVFNQAGLANILGEV 1881
Qy 254 -----MGNV---WMGRLOVYAYLAPSYSTINTSKVTGEVFNHLLTVGDK--NAA 298
Db 1882 AVQSINKAGGLNGLNLTVLTGNSVIGGYLTPQKQNTLSQLLQGNFNDNL--MNDSGLNTA 1940
Qy 299 QAGLIANKNTNIGTLDLWQS-----AGLNIITAPPEGGYKDKPNTTPSOGAKNDKNE 350
Db 1941 IKDLIRQK-----LGFMTGLVGLAGLGGIDIQNPEKLIGMSINDLLSKGLFNQITG 1994
Qy 351 --SAKNDQESSNNSNTQVNPSPNAOKTEVQP--TQVIDGPFAGGKDTVNTI----- 400
Db 1995 FISANDIGQVTS---VMLQDIVKPSNALKNDVYALGQMI-GEFL-QGDTLNSLESLLQN 2049
Qy 401 NRINTNADGTIRVGGFKASLTNAAHL--HIGKGV-----NLSNQASGRSL 445
Db 2050 QOIKSVLDKVLAAKGLPIYEQGLGLIPNLGKGLFAPYGLSQVQKGFSPNAQG--NV 2108
Qy 446 IVENLTGNTIVDGLRVNNOVGGYALAGSSANPEFKAGTDTKNGTATFENDISLGRFVNL 505

Db 2109 FVQNSTSNANGGTLSFN-----AGNSLIF---AG-----NNHIA---FTN- 2143
Qy 506 KVDATHTAFKGDITGNGGFENTLDFSGVTDKVNINKLITASTNVAVKFNFI----- 555
Db 2144 --HAGTQLQLSDQVSNINITLNAS-----NGLKIANNVSVSOGNLFVSASCAQOS 2195
Qy 556 -----NELIV---KTNGISVGEYTHFSDIGSQRINTVRLETGTRSL-----F 596
Db 2196 DPTTANIANPCALSAQSTNGAS-----SNASNAPIALSNDESLSMAVANDFN 2245
Qy 597 SGGVKFGGKGLYDIEFYSPWNYFDARNIKVBEITNKLAFGPOGSPWGTSKLMFNLT 656
Db 2246 SGNLYANG---VVD---FS--KIKGSANIKNLYLNAQF-----QANNLT 2284
Qy 657 GQNAVMDYSO---FSLNLTIOGDFINNOGTINYLVRGKVATLSVGNAAAMFNNDIOSAT 713
Db 2285 SNQAVLEKNASFVNNLNIOGAFNN-----ATQKIEVLQNLVIASNASLST 2331
Qy 714 GFYKPLI--KINSADLIKNTHEVLLKAKII---GYGNVSTGTNGISNVNLEEQKERLA 768
Db 2332 GIYGLEVGALNNSGAIFHNLENTQTPTLQAEGIIINLNTQTTPFMNVN----- 2381
Qy 769 LYNNNNMDTCVVRNTDDIKACGMAIGDQSMVNNPDNPKYKILGAKWNIGISKITANGSKI 828
Db 2382 -----NSMANN--TTTTLKSSRYIDYNINPNSLOSYL 2412
Qy 829 SVYYL---GNSTPTENGNT-----TNLPTNTTSNARSANNAALQNA- 867
Db 2413 NLYTLININGNHIEEKNGALTYLQORVLLQDKGLLSVALP---NSNASONNLSLSVL 2469
Qy 868 -----PFAQPSATPNLVAI--NOHDFGTIESVPFELANRSKIDIDFLYANSQAQ 913
Db 2470 YNOVKMSGCGKAMDFTPTLQDYIVGIGQSALNQIEAV-----GGNA 2512
Qy 914 RDLQTLILDS--HDAGYARKMIDATSANEI---TKQLNTATTTLNNTASLEHTSGLOTL 969
Db 2513 IKWLSTLMMETKPPFPFIYLNKHSLEILGVTKDLQNTASLNSPNFRDNATNLLELA 2572
Qy 970 SLSNAMILNSRLVNLRRHTNHIDSFAKRLQALKDQKFALESAAEVLQYQAPKYEKPTN 1029
Db 2573 SYTQ---TSLTKLSDFRSREGESFSLLE-LKNKFS--DPNPEVYVYQSLSKHPPN 2626
Qy 1030 VWANAIGTSLNNGSNASLYGTSAGVDAYLNGQVEAIVGGFSGYSGYSFNNRANSLANGA 1089
Db 2627 LHWQGVGGSFISGNGTLYGLNAGYDLRVK---NVILGGYVAYGYSDFN--GNIMHSLG 2681
Qy 1090 NNTNFGYVSRIFANQHEFEAOGALGSDQSILNFKSALLQDLNOSYHVLAYSAAATRAS 1149
Db 2682 NNVDMGYARAFKRNFTLSANEYGGNATSNSSLSLSVLNQRYNTWTTSVNGNY 2741
Qy 1150 GYDPAFFRNALVLPKPSVGSYNHLGNTNPKSNST--NQVALKNGSSSQHLEFNASANVEAR 1207
Db 2742 GYDPMFKQKSVVLKPOVGLSYHFGLSGMKGNDAAAYKQFLMHSNPSNESVLTLMNGLES 2801
Qy 1208 YYGDTSYFYNNAGVQLQEF--FAHVGSNNAASL--NTEFVNAARNPLNTHARVMGGEUKL 1263
Db 2802 KYFGKNSYVFTVTLARGLDLILAKAGDNVVRFGENTL-----LYRKGIFN 2856
Qy 1264 AKEVFLNGLVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1296
Db 2861 WRLVYVNAVGLKMGLOQYQDINTGNGMRVAF 2893

RESULT 8

B64635

toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pyloriC:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: B64635R:Tomb, J.F.; White, O.; Kervage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A:Reference number: A64520; MUID:97394467
A:Accession: B64635
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2529 <YOM>
A:Cross-references: GB:AE000602; GB:AE000511; NID:g2314060; PIDN:AAD07969.1; PID:g231406

Query Match 9.6%; Score 642.5; DB 2; Length 2529;
Best Local Similarity 23.0%; Pred. No. 1.2e-24;
Matches 294; Conservative 188; Mismatches 496; Indels 301; Gaps 50;

Qy 149 RNFTEGDLV-----NMOKATLRLGQFNGNSFTS---YKDSAD-----RTT-----186
Db 1421 OFENGNLIIGATQDNVTNSTSSI-WFGNGYSSPCTLDSATCSFRNTYLQQLGSTSP 1479
Qy 187 -----RVDFNAKNIIDNFVEINNRVGSAG-----RKASSTVLT-----LQASEGI---228
Db 1480 YLGYINADFKAKSIYITGTIGSGNAFEGSGSADVTFSANNLVNLANIEAQAOTDNIFNL 1539
229 TSDKNAEISLYDGATLN---LASSSVKL---MGVNMGRLOYVGCAYLAPSYSTINTSKV 281
Db 1540 LQKGIEKIFNOGNLANVLSQVAMEKIKQAGGLGFENALSPLSKEUPASLQNETLQOL 1599
Qy 282 TGEVNFNHL--TVGDKNAQAAGIIANKTNICTLQWSAGLNIIAPPBGYKDKPNNTP 339
Db 1600 IQGNLDDLNNSGVMNAIQ-NIISKLSIFG-----NFV-----TP 1635
Qy 340 S--QSGAKNDKNESAKNDK-----OESSQNSNTQVI-----NPPNSAQKTEVOPT 383
Db 1636 STIENYLAQSKLSMLDKGLLNFIFGYMNASELSSILSVLKDITNPTSLQKDIGVVA 1695
Qy 384 QVIDPFAAGKDTVINRINTNADGTRVGGFKASLTNAHLHIGKGVN-LSNQASG 442
Db 1696 NDLLNEFL-GQDVIKKE-----SGLVSNIIINIIISOGGLSGVYNQGLG 1739
Qy 443 -----RSLIVENITG-----NITVDGRLRVNNOGVGYALAGSSANF 478
Db 1740 SVLPSPQLNAKENDGLTLLSPRGLHDFWOKGYFNLSNGYVFNSSFSNA-TGSLNF 1798
Qy 479 EPKACTDFTKGTATFNNDISLGRFVNLKVDAHTANFKGIDTGGNGFNTLDFS-----530
Db 1799 -----VANKSIIFNGD-----NTIDFSKYQCALIF 1823
Qy 531 GVTDKVNTKLITATNVAVKNFNNELIVKTINGISVG-----EYTHFSEDIGSQSR 582
Db 1824 ASNDSVINITTLNATNGLSLNLNNSVQKGEICVNLNCPPTKNSSTSSVTPTNE 1883
583 INTVLETGTRSLFSGGKFKGGEKL-----VIDEFYSPWNYFDARNIKNVEITNKL 635
Db 1884 SLSVRANFT---FLGATASNGAIDLSQVKNNSVIDTLNL---NENAAQLANNLTIN-- 1935
Qy 636 AFGPGSGPWGTSKLFNNLTGQNAVMDYSQPSNLTIGDFTNNOGTINY-LVRGKGVAT 694
Db 1936 AFNNASN---STANINGNFTLNQATLS--TNASGLNVMGNF-NSYGDVLVNLSSHVSVAI 1990
Qy 695 LSVGNAAAMFNNDIDSATGFKYKPLIKNSAQDLIKNTEHLLKAKII--GYGNVSTGTN 752
Db 1991 NAAQSATIMANN-----NPLQIOTNKSKEV-GTYTLIDSAKAIYGYNNQITGS 2041
Qy 753 GISNVNLEEQFERKALYNNNNRMDTCVVRNTDDIKACGMAIGDSQSMVNNPDNYKILGK 812
Db 2042 SLDNY-----LKLY-----TLI-----DINGRHVMTDNGL-----2067
Qy 813 AKWNGISKTANGKISIVYLGNSPTPTGGNTTNLTPTNTTSNARSNANALQAONPAPAP 872
Db 2068 -----TYNGQAVS-----KDGGLVWGFKD-----SQNGYIVTSILYNKV 2102
Qy 873 SATPNLVAIQHDFCTIESVFELANRSKDIDFLYANSAGQGRDILLQTLIDSHDGYARK 932
Db 2103 KTAVSNDPNNILQAPTLKQYIAQIOGTGVDSDIOAGGSAQINLNLNFKETKGPLFAPY 2162

Qy 933 MIDATSAANEITKQLNTATTTLNNIASLEHKTSLGQTLSSLNAMIILNSRLVNLRRHTNHI 992
Db 2163 YLESHSTKDLTTIAGDIANTLEVIANPNFKNDATNLIQINTYTOQMSRLAKLSDSTFPAS 2222
Qy 993 DSFAKRLQAKDQKFA-SLESAAEVLYQAPKYEKPTNVWANAIGGTSLNNGSNASLYGT 1051
Db 2223 ADHERLEALKRKRFADATPNDAMDVLKYSQRNRYKNNVWATGVGGASPTGCTGLYGI 2282
Qy 1052 SAGVDAYLINGQVEAJVGFGSYGSFENNRANSLANGSANTNFGVYSRIFANQHEPDFEA 1111
Db 2283 NVGYDRFNG---VIVGGTAAYISGFH--ANITQSGSSNVNMGVYSRAFIRKSEUTMSL 2337
Qy 1112 QGALGSDQSSALFKSALLQDLNQSHYLAASAATRSYGYDFAFFRNALVLKPSVGSYN 1171
Db 2338 NETWYKNTFINSYDPLLSIIINQSYKYDTWTWDKINQYDFEMDKSVIFEPQIGLAYY 2397
Qy 1172 HLGSTNFKNST---NOVALKNGSSQHLFNASANVEARYYYGDSYFYMAGVLOEF- 1226
Db 2398 YIGLSGLRGIMDPDPIYNQFRANADPNKKSVLTINFALSHRYFNKNSYYFEYFVADVGRDLF 2457
Qy 1227 -----AHVGSNNAASLNTFKVNAARNPLNTHARVMGMGELKLAKVEFLNLGVVYLH 1277
Db 2458 INSMGDKMVRFGNNTFLSYROGGR-----YNTFASIITGGEIRLFTFYVNAIGARF 2510
Qy 1278 NLISNIGHFASNLGMRYSF 1296
Db 2511 GLDYKDNITGIMGRYAF 2529
RESULT 9
H71879
toxin-like outer membrane protein jhp0856 - *Helicobacter pylori* (strain J99)
C:Species: *Helicobacter pylori*
A:Variety: strain J99
C:Date: 12-Feb-1999 sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: H71879
R:Alm, R.A.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: H71879
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2399 <ARN>
A:Cross-references: GB:AE001515; GB:AE001439; NID:g4155425; PIDN:AD06432.1; PID:g415
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0856

Query Match 9.3%; Score 623; DB 2; Length 2399;
Best Local Similarity 22.7%; Pred. No. 1.1e-23;
Matches 296; Conservative 175; Mismatches 479; Indels 356; Gaps 51;

Qy 149 RNFTEGDLV-----NMOKATLRLGQFNGNSFTS---YKDSAD-----RTT-----186
Db 1292 OFENGNLIIGATQDNVTNSTSSI-WFGNGYSSPCTLDSATCSFRNTYLQQLGSTSP 1350
Qy 187 -----RVDFNAKNIIDNFVEINNRVGSAG-----RKASSTVLT 222
Db 1351 YLGYINADFKAKSIYITGTIGSGNAFEGSGSADVTFSANNLVNLANIEAQAOTDNIFNL 1410
Qy 223 QASEGITSKNAEISLYDGATNLNASSSVKL---MGVNMGRLOYVGCAYLAPSYSTINTS 279
Db 1411 LQSEGI--DKIFNQGNLANVLSQVAMEKIKQAGGLGFENALSPLSKEUPASLQDETGL 1468
Qy 280 KVTGEVNFNHL--TVGDKNAQAAGIIANKTNIIGTL-----DLWQSAG- 320
Db 1469 QLIGQNLLDDLNNSGVMNEIQ-NIISQKLSIFGFNFTSIENYLAQSLAKSLMLDKGL 1527
Qy 321 LNIIAPPEGYKDKPNNTPSQSGAKNDKNESAKNDKQESSQNSNTQVI-----NPPNSA 375

Db 1528 LNFI-----GGVID-----: : : : : ||| |
Qy 376 QKTEVQPTQVTDGPFAGGKDVVNINRINTNADGTIRVGGFKASLTNTAAHLHGKGGVN 435
Db 1559 QKDGTVVANDLNEFL-GQDVVKLE-----SOGLVSNIIINNVISOGGLS 1602
Qy 436 -LSNQASG-----RSLIVENLGT-----NITVDGPLRVNQGGYA 470
Db 1603 GVINOGGLSVLPSSLQALKENDGLTLLSPRGLHDFWQKGYFNLNSGYFVFNSSFSNA 1662
Qy 471 LAGSANPEFKAGTDTKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGGNGGNTLDFS 530
Db 1663 -TGGSNAF-----VANKSIIFNGD-----NTIDFS 1686
Qy 531 GVTDKVNINKLITASTNVAVKNFENELIVKTINGISVGEYTHFSEDIGSQSRINTVLEET 590
Db 1687 KYQ-----GALIFASNGVS--NINITLNA--TNGLSL-----NAGLNVSVOK 1726
591 G-----TRSL-----FSGGVRFKGGKL--VIDEYY 615
Db 1727 GEICINLANCPTTKNSSPANSVPTNESLSVHANNFTFLGTIISNGAIDLSQVTNNSVI 1786
Qy 616 SPWNYFARNIK--NVEITNKLATGPOGSPWGTSKLMPNNTLQONAVMDYQSFNSLTIQ 673
Db 1787 GTLNLNENATLQANNLITN--AFENNASN--STANIDGNFTLQOATLS--TNASGLNVM 1840
Qy 674 GDFTNNOCTINY-LVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLIKINSADQLKNT 732
Db 1841 GNF-NSVGDVFNLSHVSIAIITQGTATIMANN-----PLIQFNASKEV--GT 1889
Qy 733 EHVLLKAKII--GYGNSTGTNGISNVNLEBQFRERLALYNNNRMDTCVVRNTDDIKAC 790
Db 1890 YTLIDSAKAIYYGNQITGSSLDNY-----LKLX-----ALIDNGK 1928
Qy 791 GMATGDOSWNNPNKYLIKAKNKIGISTANGSKISYYLGNSTPTENGNTNLTPT 850
Db 1929 HMVMTDNGL-----TYNGQAVS-----KDGGLVWGFKD 1957
Qy 851 NT-----TSNARSANALAQNAPEAQSATPNLVAIQHDFGTIESPELANRSKIDITL 905
Db 1958 SONQYIYTSILYNKVKTAVSNDPINNQPAP-----TLKQYIAQGVQSVDSI 2005
Qy 906 YANSAGQGRDLLQTLIDSHDAGYARKMIDATSANEITQLNTATTTLNNTIASLEHKTSG 965
Db 2006 DQAGNOAINLNLKIFETKGPLFAPYLYESHSTKDLTTIAGDIANTLEVIANPNFKNDA 2065
Qy 966 LQTLSLSNAMTLNRLNLSRRHTNHIDSFAKRLQALKDKQEA-SLESAAEVLQFAPKY 1024
Db 2066 TNLQINTYTOOMSRKLKSDTSTFARSDFLERLEALKNKRFAADAI PNAMDVILKYSORN 2125
Qy 1025 EKPTNVMANAIGGTSLNNGSNASLYGTSGAVDYLNGQVEAIVGFGSGYGYSSFFNNRANS 1084
Db 2126 RVKNVWATVGGASFTSGGTGLYINGVYDRFIKG---VIVGGYAYGSGFH--ANI 2180
Qy 1085 LNSGANNTNFVGYGRIFANQHEFDEAQAALGSDOSSLNFKSALLQDLNOSHYLAYSA 1144
Db 2181 TQSGSSNVNVGVYRAFTRKSELMTSLNETWGNKTFINSYDPLLSIINQSYRYDTWTD 2240
Qy 1145 TRASYGVDFAFRNALVLKPSGVSYNHLGSTNFKSNST-----NOVALKNQSSSQHLENA 1200
Db 2241 AKINGYDFMEFKDSVIFKPOVGLSYIYIGLUGLGRIMDDPIYNOFRANADPNKKSVLTI 2300
Qy 1201 SANVEARYYYGDTSYFYNNAGVLOEF-----AHVGSNNAASLNTFKVNAARNPLN 1250
Db 2301 NFALESRYHFNKNSYFYVIADVGRLDFINSMDKWKVRFIGNNTLSYRDGGR-----YN 2353
Qy 1251 THARVMGGEKLAKAEVPLNGLVGYLNLHLSIGNIGHFASNLGMYSF 1296
Db 2354 TFAIITGGEIRLFKTFVNNAGIGARFGLDYKDINITGNIGMRVAF 2399

RESULT 10

B64596
toxlin-like outer membrane protein HP0610 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: B64596
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, R.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: B64596
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-1943 <TOM>
A:Cross-references: GB:AE000575; GB:AE000511; NID:q2313730; PIDN:AAD07675.1; PID:q231
C:Genetics:
A:Start codon: GTG

Query Match 8.7% Score 583: DB 2: Length 1943;
Best Local Similarity 22.3% Pred. No. 9.2e-22;
Matches 330; Conservative 212; Mismatches 540; Indels 398; Gaps 70;
Qy 6 THRKINRPLVSLALVGALVSITPQO-SHAFFTTVIIPAIVGGIATG-----TAVGT 56
Db 673 TYNAGNQPLQALHIYNQ--AITKODLNMIASLGKEFLPKIANLLSSGALDNLNPSNSET 730
Qy 57 VSGLL-SWGLLQKQAEANKTPDKPKVWRIQACKGFNEFPN-KEYDLYRSLSSKIDGWD 114
Db 731 LFGIFERYGITLNGEN-----WK-SLLKIIINFNTNTNYDFSQGLN---VWGA-- 774
Qy 115 WGNARHWYKGGQONKLEV-----DMKD--AVGTYTSLGRNFTGGDLVDVNMOKATLR 166
Db 775 -----IKEGDTNKTYSVWFGGEGYKEPCAVDNTQCMFROTNLGQL---LHSSTPY 822
Qy 167 LGQFNGNSFTSYKDSADRTTRVDNFNAKNISIDNEFINNRVYSGAGRKAS-----STVLT 221
Db 823 LGINAN-----FRAKNIYITGTIGSGNANGSGGSANVSFESGTLNLVLN 866
Qy 222 LQASEGITSOK-----NA-----EISLYDGA-TLNLAASSVKLM 254
Db 867 QAKIDAGOTDKIFSYLGQGGIEKLFEGKGLGNALSNIIYESLNDNAIPKOLANMIPKDF 926
Qy 255 GNVWNGRLQYVGAYLAPSYSTINTSKVTGEVNFNHLTVGD-KNAQAAGIIANKKTNIGTL 313
Db 927 GSKTSSL-----LSPT-----EYN-NLLGVSAFKNAIME--ILNSKT-VG-- 963
Qy 314 DLWQSAG-LNTIAPPE-----GGYK-----DKPNNTPSQSGAKND 347
Db 964 DVFGENGLLNALDPTERRKIDOMLLEQIOAHSSGFEKFIKVTGLTIENVENFINNWKQS 1023
Qy 348 KNESAKN-----DKQESSONNSNTQVINPNSAKQT---EVQPTQVIDGP----- 389
Db 1024 LSSFANNFVPGGLNQALDKIGSSSDAKDLQ-----NFDLTKTTFGDILNQMTQEQAPL 1078
Qy 390 --FAGGQDVTVIN-RINTNADGTRVGGFKASLTNTAAHLHGKGGVN--LSNOASGR- 443
Db 1079 ISWLGPDQLSVLVNIALNSITNPSKELTSTISSIGERKALNDLLGDGVVKNKTSNQVLQOM 1138
Qy 444 --SLIVENLTGNITVDG-----PLRVNNQ-----VGGYALAGS-----SANFEKAGTD 485
Db 1139 INKTIADKFGGVYQOGLGSLTPOSLODELKKGWGLSGSLGRGLNHLNWRGNFNFVAKDY 1198
Qy 486 TKNTATFNNDISLGRFVNLKVDHAHTANFKGIDT-----GNGGFNTLDSGYT-DKVN 537
Db 1199 LFTNNSFSFN--ATGGELNF-VAKGSIIFNGKNTINTFYQOGKLSFISKDFSNISLOTLN 1255
Qy 538 INKLITASTNVAVKNFENELIVKTINGISVGEYTHFSEDIGSQSRINTVLEET---GTR 593
Db 1256 -----ATNGLTLNAPKNDISVQGCIVNVNLCMGEKKHAHSSSATAPTNETLEANANN 1308

QY 832 YLGNSTPTENGNTNLTPTNTTSNARSANALAQAPFAQSPATPNLVAINHDFGTIES 891
 Db 2810 WLSELFPAK--GGNPLFAPY-----YLDQNPTEHIVL----- 2840
 QY 892 VFELANSKIDIT---LYANSAGGR--DLLQTLILDSHDAGYARKMIDATSAEITKOL 946
 Db 2841 -----MKDITTSALGMLSNLKNSTDLV----- 2866
 QY 947 NTATTTLNNTASLEHKTSGIQTLSLNSAMILNSRLVNLRRHTNHIDSFARLQALQKOK 1006
 Db 2867 NTTYQOMSLRLAKLSNFASFST-----DFSERLSLKNQR 2901
 QY 1007 FA-SLESAAEVLQFAPKYBKPTNVWANAIGGTSNLNNGSNASYGTSAGVDAYLNGOVEA 1065
 Db 2902 FADAVPNAMDVILKYSQDKLKNLWATGVGVSVFENGTTGLYGVNVGVDRFVRG---V 2958
 QY 1066 IVGGFGSYGSSFNRRANSLSNGANTNFVGYSRIFANQHEFDEFAOGALGDSQSSLNFK 1125
 Db 2959 IVGGYAAAGYSGFYERTS--SKSDNDVDVGLYARAFIKKSELTFVSNETWGANKTISN 3016
 QY 1126 SALLQDLNQSHYLAASAATRASGYDFAFRNALVLKPSGVSYNHLGSGTNFKS--NST 1183
 Db 3017 DALLSMLNQSYKYSTWTTNAKVNGYDMEKKNKSIILKPOIGLRYYYIGHSGLEGVMNV 3076
 QY 1184 --NOVALKNGSSOHLFNASANVEARYYYGDTSYFYNNAGVLQBEFA--HYGSNNAASLNT 1239
 Db 3077 LYNQFANADPSKSVLTIDFALENRRHYFNTSYFYAIGGVGRDLLVNSMGDKLVREIGN 3136
 QY 1240 FKYNARNPL-NTHARYVMGELKLAKEVFLNLGVVYVYLNHLNISIGHFASNLGRYSF 1296
 Db 3137 NTLSYRGDLTYNFANITTEGEVRLKFSFYANAGVGARFGLDYKMDIIGNIGMRLAF 3194
 RESULT 12
 E97835
 hypothetical protein romB [Imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
 R:Qogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
 Science 293, 2093-2098, 2001
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; PMID:2142074; PMID:11557893
 A:Accession: E97835
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1655 <KUR>
 A:Cross-references: GB:AE006914; PIDN:AAL03623.1; PID:g15620207; GSPDB:GN00173
 C:Genetics:
 A:Gene: romB
 Query Match 5.0%; Score 335; DB 2; Length 1655;
 Best Local Similarity 19.5%; Pred. No. 4e-09;
 Matches 274; Conservative 177; Mismatches 531; Indels 420; Gaps 59;
 QY 146 SGLRNFTGGDLDMQK-ATLRLQFNGNSFTSYKDSADRTTRVDENAKNISIDNFEIN 204
 Db 211 NGTLNVTNGFIKYSKSFATVNVINIGDGGIMNTDADNVNTLNQANGATI-----TFN 266
 QY 205 NRVGSG-----AGRKASSTVTLQAS-----EGI-----TSDKNAEISLYDGA-----TLN 245
 Db 267 GTDGTGRLVLKNAATDFNVTSGLGNLKIIEFTVAVNGOLKANAGANAAGVICTNN 326
 QY 246 LASSSVLMGNVMGRLOYVCVILAPSYSTINTSKVTGEVNFHL-----TVGDKNAA 298
 Db 327 GAGRAAGFVSVNDKGVATIDGGVYAKDVIQSANAGVQNVFRHIVDVGDTGTFATKTA 386
 QY 299 QAGIIANKKTNIGTLDLWQSGAGNIIATPE-----GGYKDKPNNTP-----SQS 342
 Db 387 -SKVAITQNSNFGTDFGNLAA-QIIVPNTMTLNGNFTGDASNPGNTAGVITFDANGTLA 444

QY 343 GAKNDKNESAKNKQESSQNNSTQVNPNSAQ-----KTEVQPTQ 384
 Db 445 SASADANAVVNTNITATEASGAGVQVLSGTHAAELRLGNAGSVFKLADGTGVINGKVNQTA 504
 QY 385 VIDGPPAGGADTV--VNINRINTNADGTTIRVGCF-----KASLTNAAHLHIGKGSVN 435
 Db 505 LVGALAAGTITLDGSTATITGDIAGNAGAAALOGIITLANDATKTLTLLGGANI-IGANGGT 563
 QY 436 LSNQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF-----KAGTDTKNG- 489
 Db 564 INFQANGTKLSTQNNIVVDFDLATATDOTGVVDASSLTNAQTLINGKIGTGVANNK 623
 QY 490 -----TATFNNDISLGRFV-----NKKVD 508
 Db 624 TLGQFNIGSSKTVLSDGDVAINELVIGNGAVOFAHNTYLTITRTNAAOGKIFNPVFN 683
 QY 509 AHTANFKGIDTGAG-----GENTLDFSGVTDKVNINKLITA-STNVAVKFNINELIV 560
 Db 684 NNTLTATGTLGSAATNPLAEINFGSKGAANVDVNLVNGKVNLYATNITTTDANGVSFIF 743
 QY 561 KTNGISVGEYTHFSEDIQSO--SRINTVRLTETGTRSLFSGGVKFKGGEKLVIDEFYFSPW 618
 Db 744 NAGGTNI-----VSGTVGGQGNKFNVALDNGTTVKFLGNATFNNGNTTIAANSTLQIGG 798
 QY 619 NY-----PDARNIKNWEITN---KLAFGPQOSP-----WCTSKLMPFNLTQ 658
 Db 799 NYTADFEVASADGTGI--VEFVNTGPTITVLNKOAAAPNALKQITVSGPGNVVINEIGNAG 856
 QY 659 N---AVMDYSQFENLTIQDGINNOGTYNLVRGKVATLSVGNAAAMFNNDI--DSAT 713
 Db 857 NYGCAVTDITAFENSSL-----GAVFELPRG--IPFNDAGNRIPLTTKSTVGNKTIAT 906
 QY 714 GFYKPLIKINSADOLIKNTEHVLKAKIIGY-----NVSTGTNGISNVNLEB--- 761
 Db 907 GFDVPSVILGVDSVIADGVIGDQNNIVGLGSDNDIIVNATTLVAGICTINNQGTV 966
 QY 762 -----QFKERLAL--YNN-----NNRMDTCVVRWTDIKA 789
 Db 967 TLGGGIPNTPCTVYGLGTGICASKFKQVTTTDDYNLGNIIATNATINDGTVTTTGGIAG 1026
 QY 790 CG-----MAIGDQSMVNNPDNYKY---LIGKAWKNIGISKANGSKISVYVLGNS----- 836
 Db 1027 IGFDKITLGS---VNGNGNRFVVDGLSHSTSMIGTTKANNG---TVTYLGNAFVGNIG 1080
 QY 837 ---TPTEN-----GQNTT-----NLPT 850
 Db 1081 DSOTPVASVRFSGDGAGLQGNISQVIDFGTYNLGINSNVILGGTGTAINCKINLRT 1140
 QY 851 NNTSNARSANALAQAPFAQSPATPNLVAINHDFGTIESVPELANRS----- 899
 Db 1141 NTLTEA-SGTSTWGNNT-----SIETTLTLANGNIGNIVILEGAQ 1179
 QY 900 -----KIDITLYAN-SGAQGRDLQ-----TLLIDSDHAGYARKMID 935
 Db 1180 VNATTCTTTIKVODNANANFSGTQVTLTQGGARFNGTLGGPNFVTVGSRNRFVNYGLIR 1239
 QY 936 ATSAEITTKOLNTA-----TTTLNNTIASLEHKTSGIQTLSLS 972
 Db 1240 AANQDYVITRTNAENVVNTDIANSFSGGAPGVQNVTFVFNATNTAAYNLLAKKNSAN 1299
 QY 973 NAMILNSRLVNLRRHTNHIDSFARLQALQKOKFASL----- 1010
 Db 1300 SANFVGAIVTDTSAAITNAQLDAKDIQAOLGNRLGALRYLGTPEAEMAGPEAGAIPTAA 1359
 QY 1011 -----ESAAEVLQFAPKYKPK---TNVWANAIGTSLNNGSNASLYCAGVDAYLNGOV 1063
 Db 1360 VAAGDEADVNAVGI---WAKPFYTDHQSCKGGLA---CYKAKTTGVVIGLDTLANDNL 1413
 QY 1064 EATVGGSGSYGSSFNRRANSLSNG-----ANNTNFGVYSRIFANQHEFDEFAOG--ALGSD 1118
 Db 1414 --MIG--AAGITKTDIKHQDYKKDKTDVNGFSFSLYG---AQQLVKNEFAOGSAIFSL 1466
 QY 1119 QSSLNFKSALLQDLNQSY-----HYLAYSAATRASGYDFAFRNALVLKPSGVGSYN 1171

Db 1467 NOVKKSORFFDANGNNKSKQIAAGHYDNMTFGNLTGVGDYDAMQGVLLV-TPMAGLSYL 1525
QY 1172 HLGSTNEKSNSTNOVALKNGSS 1193
Db 1526 KSSDENYKETGTT-VANKOVNS 1546
RESULT 13
S07575
outer membrane protein B - Rickettsia rickettsii
N:Alternate names: 120K surface-exposed-protein
N:Contains: 32K beta peptide
C:Species: Rickettsia rickettsii
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Dec-1998
C:Accession: S07575; S18579; A43869
R:Gilmore Jr., R.D.; Joste, N.; McDonald, G.A.
Mol. Microbiol. 3, 1579-1586, 1989
A:Title: Cloning, expression and sequence analysis of the gene encoding the 120kD surface
A:Reference number: S07575; MUID:90136087
A:Accession: S07575
Molecule type: DNA
Residues: 1-994, 'DLKLEH', 1001, 'GS', 1005-1300 <GIL>
Cross-references: EMBL:X16353
A:Note: This sequence has been revised in reference S18579
R:Gilmore, R.D.; Joste, N.; McDonald, G.A.
Mol. Microbiol. 5, 3089, 1991
A:Reference number: S18579; MUID:92236427
A:Contents: erratum
A:Accession: S18579
A:Molecule type: mRNA
A:Residues: 994-1004 <GIL2>
R:Hackstadt, T.; Messer, R.; Cieplak, W.; Peacock, M.G.
Infect. Immun. 60, 159-165, 1992
A:Title: Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein
A:Reference number: A43869; MUID:92104668
A:Accession: A43869
A:Molecule type: protein
A:Residues: 1008-1027 <HAC>
A:Note: sequence extracted from NCBI backbone (NCBIP:74388) and corrected to correspond
C:Keywords: glycoprotein
F:2-1007/Product: outer membrane protein B #status predicted <MAYO>
F:1008-1300/Product: 32K beta peptide #status experimental <MATB>
Query Match 4.7%; Score 314.5; DB 2: Length 1300;
Best Local Similarity 21.0%; Pred. No. 3.2e-08;
Matches 274; Conservative 164; Mismatches 489; Indels 377; Gaps 60;
QY 111 GWDMGNAARHYVVRGGQONKLEVDKDAVGYTTLISGLRNFYGGDLVDNMQRATLRIGQF 170
44 GNTDFGNLA-----AQIKVPNAI---TLTG--NFTG---DASNPGNTAGVITF 83
QY 171 NGNSFTSYKDSADRTVRDFNAKNISIDNFVNNRVSGAGRKASSVTTLTQASEGITS 230
Db 84 DANG-TLESAD-----ANVAVTNITAIASGAGVYQLSGT-----HAAELRLG 128
QY 231 DKNAEISLYDGATLNLASSVKVLMGNVGRQLQYQYQYAYLAPSYTINTSKVTGEVNFNHL 290
Db 129 NAGSIFKLDAGVINGKVNQTPALVG-----GALAAGTITLDGSATITGDIG---- 174
QY 291 TVGDKNAQAAGIAIANKKTNIGTLDLWQASAGLNIIAPPEGGYKDKPNTPSQSGAKNDKNE 350
Db 175 NAGGAALQRITLANDAKKTLT-----CGANII-----GAGGGTIDLOA 214
QY 351 SAKNDKQESSQNNSTQVINPNNSAQKTEVQPTQVIDGPPFAGKDTQVNVNINRINTNADGT 410
Db 215 NGGTIKLTSTQNNI-----VVDFDLA-----IATDQTCV 243
QY 411 IRVGGFKALTTNAHLHI-GR-GGVNLSNQASGRSLIVENTGNITVDGPLRVNNOVGG 468
Db 244 V-----DASSLNTAOTLTINGKIGTGIGANNKTLGOFNIGSSKT--VLSNGNVAINELVIG 296

QY 469 YALAGSSANFEFKAGDTQTKNGTA-----TFNNDISLGRFVNKLKVDARTAFKIGDIRG 520
Db 297 NDGAQVFAHDTYLI-TRTNAAAGQGLIIFNPVYNNGTTLAAGTNL--GSATNPPLAEINFG 353
QY 521 NCGFWTLDFSGVTDKVNINKLITASTNVAKNFNINELIVKTNISVGEYTHFSEIDIGSQ 580
Db 354 SKGVNVDTVLVNVEGVNL-----YATNITTTDANVGSFVNAGGTNI-----VSGTVGGQ 403
QY 581 --SRINTVRLGTGTRSLFSGGVKFKGGEKLVIDEFYSPWNYDARNIKNVEITNKLAF- 637
Db 404 QGNKFTVALENGTTVKFLGNATFNGNTTIAANSTLIQICGNY-TADCVASADGTGIVEFV 462
QY 638 --GP-----OQSPMGTSKLMFNNTLQON---AVMDYSQFSNLTIQGDF 676
Db 463 NTGPITVTLNKEAAPVNALKQITVSGPGNVVINEIGNAGNHHGAVTDTTAFENSSL---- 518
QY 677 INNOGTINYLVRGGKATVLSVGNAAAMFNNDIDSAT--GFYKPLIKINSADOLIKNTEH 734
Db 519 ----GAVVFLPRG--IPENDAGNTMPLTIKSTVGKNTAKGFDVPSVVLGVDSVIADGQV 572
QY 735 VLLKAKIIGYG-----NVSTGTNGISNVNLEP----- 761
Db 573 IVDQNNIVGLGSDNGIIVNATTLYAGISTLNNNGTVTLSGGVPNTPGTVYGLGTGIG 632
QY 762 --QFKERLAL--YNN-----NNRMDTCVVRNTDDIKACG---MAIGDQSMVNNPDNY 806
Db 633 ASKEKQVTFPTDYNLGNLIATNATINDGVTVTGTGGIAGIGFDGKITLGS---VNGNGNV 689
QY 807 KY---LIGKAWKNIGISKANGSKISVYVLGNSTPTENGCGNTNLP---TWNTTSNARSA 859
Db 690 READGILSNSTSMIGTTRKANG---TVTYLGNFVGNIGSDTTPVASVRFTGSDSGAGLQ 746
QY 860 NNALAGNAPFAQPSATPNLVAIHOHF-----GT----- 888
Db 747 GNIYSQVIDF---GTYNLGIIVNSNIIILGGGTAINGKIDLVNTLTTFASGTSTWGNNTS 802
QY 889 IESVFELANRS-----KIDITLYAN-SCAQRDLIQ----- 918
Db 803 IETTLTLANGNIGHIVILEGAQVNTTTTGTITKVDNANANFSGTQTVTLIQGGARENG 862
QY 919 TLLIDSHDAGYARKMIDAT--SANE---ITKQLNTATTTLLNNIASLEH----- 961
Db 863 TLGSPNFAVTGNSRNFVNSLIIRAAQDYYVITRNAENVYVITNDIANSPEFGAPGVQDQNYT 922
QY 962 -----KTSGLQTL-----SLSNAMILNSRLNLSRRHTHSDPAKRLQALKDORFASL 1010
Db 923 TPNVNTATNAYNNLLAKNSANSFVGAIVDTTSNATNVQLDLAKDIOAQLGNRLGAL 982
QY 1011 ESAAEVLVYQFAPKYEKPTNVWNAI-----GGTSLNN----- 1042
Db 983 R-----YLGTPETAEMAGPEAGASAAVAAGDEADINVAYGIAWAKPFYVTAHQSKKGL 1036
QY 1043 -CSNASLYGTSAGVDAYLNGQVEAIVGGGSGYSGYSFNRRANSLSG-----ANNTNFGY 1097
Db 1037 AGYKAKTTGVVGLDPTLANDNL--MIG--AAIGITKTDIKHODYKKGDKTDVNGFSFSLY 1092
QY 1098 SRIFANOHEFDEPAGC-ALGSDQSSLSNPKSALLQDLNQSY-----HYLAYSAATPRSY 1149
Db 1093 G---AQQLVKNFFAOGSAIFSLNQVKNKSQRYFFDANGNHSKQIAAGHVDNMTFGNLTIV 1149
QY 1150 GYDFAFFRNALVKPSVGSYNHLGTSNFKNSNTNOVALKNGSS 1193
Db 1150 GYDYNAMQGVLV-TPMAGLSYLKSSDENYKETGTT-VANKOVNS 1191
RESULT 14
A41477
190K surface antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 08-Oct-1999
C:Accession: A41477
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.
Infect. Immun. 58, 2760-2769, 1990

A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated, ne
A:Reference number: A41477; MUID:90354033
A:Accession: A41477
A:Molecule type: DNA
A:Residues: 1-2249 <AND>
A:Cross-references: GB:M31227; NID:gl52465; PIDN:AAA36380.1; PID:gl52466
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for residue
C:Keywords: surface antigen; tandem repeat
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>

Query Match 4.7%; Score 313; DB 2; Length 2249;
Best Local Similarity 19.5%; Pred. NO. 8.4e-08;
Matches 354; Conservative 193; Mismatches 589; Indels 676; Gaps 83;

Qy	14	LVS	LALV	GV	ALV	SV	TP	QO	SH	AF	FT	T	V	-----	I	I	P	A	I	V	G	I	A	T	G	T	A	V	A	G	T	V	S	G	L	S	W	G	L	66													
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:														
	31	LSS	G	AL	GV	AT	G	V	I	A	T	N	N	A	A	F	S	N	N	G	N	N	N	N	E	I	T	A	-----	G	V	A	N	G	T	P	A	G	P	Q	N	-----	85										
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:													
	67	Q	A	E	A	N	K	T	P	D	K	V	W	-----	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:													
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:														
	86	Y	G	D	Y	T	V	T	A	D	A	R	I	I	K	A	I	N	V	A	G	T	P	V	G	L	N	I	T	Q	N	T	W	V	G	S	I	I	K	N	L	P	T	L	N	A	G	S	L	T	L	N	145
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
	91	-----	N	E	P	P	K	E	Y	D	L	R	S	L	-----	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
	146	G	N	N	A	A	N	H	G	F	D	A	P	A	D	N	Y	T	C	L	G	N	I	A	L	G	G	A	N	A	L	I	I	O	S	A	A	P	S	K	T	I	A	G	N	I	D	G	-----	202			
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
	117	--	N	A	A	R	H	Y	W	K	G	-----	G	O	O	N	K	L	E	V	D	M	K	D	A	V	T	Y	T	L	S	-----	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
	203	K	T	D	A	-----	I	N	G	T	I	G	N	A	L	-----	A	V	N	V	G	A	T	L	G	G	A	V	I	K	A	T	T	K	L	T	N	A	S	V	L	T	L	T	N	A	N	A	V	-----	257		
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
	148	----	L	R	N	F	T	G	D	-----	L	D	V	N	-----	M	O	K	A	T	L	R	L	G	O	F	N	G	S	F	T	S	Y	K	D	S	A	-----	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
		:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
	258	L	T	G	A	I	D	N	T	T	G	D	N	V	L	N	L	N	G	A	L	S	Q	V	T																												

RESULT 15

RESULT
C48399

C46399
ABC-type transport protein ydbA.2 - *Escherichia coli*

C; Species: *Escherichia coli*

C:\OPERATES\BUDGETING\COI1
C:\Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #text_change 08-Oct-1999

C; Accession: C48399; D64891; H64891

R: Moszer, I.; Glaser, P.; Danchin, A.
 Biochimie 73, 1361-1374, 1991
 A: Title: Multiple IS insertion sequences near the replication terminus in Escherichia coli
 A: Reference number: A48399; MUID: 92190338
 A: Accession: C48399
 A: Molecule type: DNA
 A: Residues: 484-2020 <MOS>
 A: Cross-references: GB:D85081; NID: g3041754
 A: Experimental source: strain K-12
 A: Note: sequence inconsistent with the nucleotide translation
 A: Note: sequence extracted from NCBI backbone (NCBI:88089, NCBI:88090)
 R: Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A: Title: The complete genome sequence of Escherichia coli K-12.
 A: Reference number: A64720; MUID: 97426617
 A: Accession: D64891
 A: Status: nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-839, 'LDLPLYFQTSVIT' <BLA1>
 A: Cross-references: GB:AF000237; GB:U00096; NID: g1787665; PID: AAC74483.1; PID: g1787668;
 A: Experimental source: strain K-12, substrain M61655
 A: Accession: H64891
 A: Status: nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 'M', 915-2020 <BLA2>
 A: Cross-references: GB:AF000237; GB:U00096; NID: g1787665; PID: AAC74487.1; PID: g1787672;
 A: Experimental source: strain K-12, substrain M61655
 C: Genetics:
 A: Gene: ydbA_2
 A: Start codon: GTG

Query Match 4.7%; Score 312.5; DB 2; Length 2020;
 Best Local Similarity 21.2%; Pred. No. 7.6e-08;
 Matches 320; Conservative 192; Mismatches 574; Indels 423; Gaps 79;

Qy 50 TGTAVGTGSLGSLGK-----QAEAKPTKPKDKVRIQAGKGFNEFPNKEDLYRS 103
 Db 165 TENADGTISLQDSNGRKATINLWQIDEANTV-----A 197

Qy 104 LLSKIDGWDG--NAARHYVKG-----GOONKLEVDKDAVGYTTLGSLRN--FTGGD 155
 Db 198 LEGVSADGATKQVYHNGELVITGDNATVNNNGKTTVPDKDSTGT--EINGNNGKVIQDGD 256

Qy 156 LDVNMKATLRIGQPNNGSFTSYKDSADR---TTRVDENAKNISIDNFEINNRVG---- 208
 Db 257 LDVS-----GGHGIDITGDSATVNDKGTMTVDPESMCIQIDGDKAIVNNEGESTI 308

Qy 209 --SGAGRKAS-----STVLTLQASEG--ITSKNAEIS--LYDGATLNL 246
 Db 309 TNGGTGTQINGDATANNNGKTTVDGKSTGTGEINGNNGKVIQDGLDVSOGGHGIDITG 368

Qy 247 ASSSVKLMGNVWMLQVYGLAPSYSTINT-----SKVTGE-----VNFNHL 290
 Db 369 DSATVDNKGTTMTVDPESIGQVQDQAVVNNEGESAITNGGTGTQINGDDATANNNGKT 428

Qy 291 TVGDKNAQAAGLIAN--KKTNIGTLDLQASAGLNLIIAPPEGYKDKPN-----NTPSOSA 344
 Db 429 TVDKDSTGTETIAGNNGKVIQDGLDVS--SGGGHGIDITGDSATVNDKGTMTVDPESIGI 487

Qy 345 KNDKNESAKNDKQESSQNSNTQV--IN-----PPNSAQKTEVQPTQVIDGPFAGGRDVTVN 399
 Db 488 QIDGQAVVNNEGESTITNGGTGTQINGDATANNNGKT-----TVDGKSTGTGKIAGN 541

Qy 400 INRINTNADGTIRVGGFKASLTNAHLHIGKGGVNLNQASGRSLIVENTLTGNITV---- 456
 Db 542 IGIV--NLDGSLTV-----TGGAH---GVENIGDNGTNNKNGDIIVVSDTSGISGLIN 588

Qy 457 -----DGLRVNNQVGGY-----ALAGSAPFEKAGTDTKNGTATFNNDLSL 499
 Db 589 GEGATVNTGDNVNSNEATGFSITNNSKVSLSAGSMQVGFSTGVDL--NGN---NNSVTL 644

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:28:49 ; Search time 18.18 Seconds

(without alignments)
2760.206 Million cell updates/sec

Title: US-09-921-157-3

Perfect score: 6671

Sequence: 1 MEIQQTHRKINRPLVSLALV.....HNLISNIGHFASNLGMRYSF 1296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

nimum DB seq length: 0

ximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match %	Length	DB ID	Description
1	6671	100.0	1296	1 VACL_HELPY	Q48247 helicobacte
2	6303.5	94.5	1287	1 VACL_HELPY	Q48245 helicobacte
3	6222.5	93.3	1291	1 VACL_HELPY	Q48258 helicobacte
4	6220	93.2	1290	1 VACL_HELPY	P55981 helicobacte
5	6108	91.6	1288	1 VACL_HELPY	Q92kw5 helicobacte
6	5237	78.5	1310	1 VACL_HELPY	Q48253 helicobacte
7	335.5	5.0	1654	1 OMPB_RICRI	Q53047 r outer mem
8	335	5.0	1655	1 OMPB_RICCN	Q9Kka3 r outer mem
9	330.5	5.0	1656	1 OMPB_RICJA	O08653 r outer mem
10	326	4.9	2003	1 YDBA_ECOLI	P33666 escherichia
11	313	4.7	2249	1 OMPA_RICRI	P15921 rickettsia
12	312.5	4.7	1300	1 120K_RICRI	P14914 rickettsia
13	310.5	4.7	2021	1 OMPA_RICCN	O52657 rickettsia
14	310	4.6	1286	1 AIDA_ECOLI	Q03155 escherichia
15	306	4.6	1643	1 OMPB_RICRP	Q53020 r outer mem
16	291.5	4.4	1953	1 BTGA_SALTY	P25927 salmonella
17	291	4.4	1645	1 OMPB_RICTY	P66989 r outer mem
18	283	4.2	1541	1 IGAL_HAEIN	P42782 haemophilus
19	280	4.2	1325	1 YDEK_ECOLI	P32051 escherichia
20	276	4.1	1545	1 IGAR3_HAEIN	P43385 haemophilus
21	265.5	4.0	1702	1 IGAR2_HAEIN	P45384 haemophilus
22	262.5	3.9	1569	1 YFJA_ECOLI	P52143 escherichia
23	261.5	3.9	1694	1 IGAO_HAEIN	P44969 haemophilus
24	253.5	3.8	1460	1 PMPC_CHLMU	Q9P1v1 chlamydia m
25	246	3.7	917	1 HXA3_HAEIN	P45355 haemophilus
26	238	3.6	1394	1 HAP_HAEIN	P45387 haemophilus
27	235.5	3.5	1276	1 PMP6_CHLPN	Q92899 chlamydia p
28	219.5	3.3	2710	1 TOXA_CLODI	P16154 clostridium
29	217.5	3.3	1341	1 VG37_RPT2	P07067 bacterioph
30	217.5	3.3	1849	1 IGA4_HAEIN	P45386 haemophilus
31	213.5	3.2	1531	1 PRPD_CHLTR	O84818 chlamydia t
32	213.5	3.2	1577	1 HLVA_PROMI	P16466 proteus mir
33	213	3.2	918	1 YWJB_CAEEL	P34487 caenorhabdi

34	212.5	3.2	1039	1 AG43_ECOLI	P39180 escherichia
35	212.5	3.2	1225	1 Y309_MYCGE	P47551 mycoplasma
36	212.5	3.2	1754	1 PMPB_CHLTR	O84418 chlamydia t
37	209.5	3.1	765	1 Y260_MYCGE	P47502 mycoplasma
38	209.5	3.1	1723	1 PM20_CHLPN	Q92812 chlamydia p
39	208.5	3.1	1532	1 IGA_NEIGO	P09790 neisseria g
40	208	3.1	2334	1 WAPA_BACSU	Q07833 bacillus su
41	206.5	3.1	928	1 HXA2_HAEIN	P45354 haemophilus
42	205.5	3.1	1770	1 PMPC_CHLTR	O84419 chlamydia t
43	205	3.1	1608	1 HLVA_SERMA	P15320 serratia ma
44	202.5	3.0	1829	1 FRPC_NEIMB	Q9JJYV neisseria m
45	201	3.0	1462	1 GTFD_STRMU	P49331 streptococc

ALIGNMENTS

RESULT	1
VACL_HELPY	
ID	VACL_HELPY
AC	Q48247: Q53434; STANDARD; PRT; 1296 AA.
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DE	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Vacuolating cytotoxin precursor.
DE	VACA.
OS	Helicobacter pylori (Campylobacter pylori).
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC	Helicobacter.
OX	NCBI_TaxID=210;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CCUG 17874 / NCTC 11638;
RX	MEDLINE=94222514; PubMed=8168917;
RA	Phadnis S.H., Ilver D.J., Jansson L., Normark S., Westblom T.U.;
RT	"Pathological significance and molecular characterization of the
RT	vacuolating toxin gene of Helicobacter pylori.";
RL	Infect. Immun. 62:1557-1565(1994).
RN	[2]
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX	MEDLINE=94216933; PubMed=8163943;
RA	Telford J.L., Ghiara P., Dell'Orco M., Comanducci M., Burroni D.,
RA	Bugnoli M., Tecce M.F., Censini S., Covacci A., Xiang Z., Papini E.,
RA	Montecucco C., Parente L., Rappuoli R.;
RT	"Gene structure of the Helicobacter pylori cytotoxin and evidence of
RT	its key role in gastric disease.";
RL	J. Exp. Med. 179:1653-1658(1994).
CC	-!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC	ULCERATION AND GASTRIC LESIONS.
CC	-!- SUBCELLULAR LOCATION: Secreted.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch)
CC	-----
DR	EMBL; U07145; AAA18867.1; -
DR	EMBL; S72494; AAB30582.1; -
DR	InterPro; IPR003842; VACA.
DR	Pfam; PF02691; VACA; 1.
KW	Cytotoxin; Toxin; Signal.
FT	SIGNAL 1 33
FT	POTENTIAL.
FT	CHAIN 34 ?
FT	PROPEP ? 1296
FT	VACUOLATING CYTOTOXIN.
FT	POTENTIAL.
SQ	SEQUENCE 1296 AA; 139760 MW; 0D1F3F71AB411447 CRC64;

Query Match 100.0%; Score 6671; DB 1; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 0; Indels 0; Gap

Qy 1 MEIOOTHRKINRPLVSLALVGVLSITPQOSHAAFFTTTIIPIAVGGTATGTAVGTVSGL 60
Db 1 MEIOOTHRKINRPLVSLALVGVLSITPQOSHAAFFTTTIIPIAVGGTATGTAVGTVSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLSLSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLSLSLSSKIDGGWDMGNAAR 120
Qy 121 HYWKGGQONKLEVDMDKAVCTYTLISGLRNFTEGGDLVNMOKATLRLGOFNGNSFTSYKD 180
Db 121 HYWKGGQONKLEVDMDKAVCTYTLISGLRNFTEGGDLVNMOKATLRLGOFNGNSFTSYKD 180
Qy 181 SADRTTRVDFNAKNISIDNFEYINNRVSGSAGRKASSTVLTLOASEGITSKNAEISLYD 240
Db 181 SADRTTRVDFNAKNISIDNFEYINNRVSGSAGRKASSTVLTLOASEGITSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWVNGRLQVYVAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQA 300
Db 241 GATNLASSSVKLMGNVWVNGRLQVYVAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQA 300
Qy 301 GIANKTNICTLWOSAGLNIITAPPEGYKDKPNNTPSOSGAKNDKESAKNDKQESS 360
Db 301 GIANKTNICTLWOSAGLNIITAPPEGYKDKPNNTPSOSGAKNDKESAKNDKQESS 360
Qy 361 QNNSNTQVNPNSAQKTEVQTVIDGPPFAGGKDTVVNINRINTNADGTTIRVGGFKASL 420
Db 361 QNNSNTQVNPNSAQKTEVQTVIDGPPFAGGKDTVVNINRINTNADGTTIRVGGFKASL 420
Qy 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNQVGYALAGSSANFEF 480
Db 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNQVGYALAGSSANFEF 480
Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFTNDFSGVTDKVNINK 540
Db 481 KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFTNDFSGVTDKVNINK 540
Qy 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIQSQRINTVRLTETRSLFSGGV 600
Db 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIQSQRINTVRLTETRSLFSGGV 600
Qy 601 KFKGGEKLVIDFYYSPNVDARINKEITNKLAFGPGSPWCTSKLMFNLLTGONA 660
Db 601 KFKGGEKLVIDFYYSPNVDARINKEITNKLAFGPGSPWCTSKLMFNLLTGONA 660
Qy 661 VMDYSQFSLTIQGDFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Db 661 VMDYSQFSLTIQGDFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Qy 721 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Qy 781 VRNTDDIKACMAIGDQSWVNPONKYLIKAKNKIGISTANGSKISVYVLGNSTPTE 840
Db 781 VRNTDDIKACMAIGDQSWVNPONKYLIKAKNKIGISTANGSKISVYVLGNSTPTE 840
Qy 841 NGGNTNLTPTTTSNARSANALAQNAFPAQSPATPNLVAIQHDFGIESVFELANRSK 900
Db 841 NGGNTNLTPTTTSNARSANALAQNAFPAQSPATPNLVAIQHDFGIESVFELANRSK 900
Qy 901 DIDTLVANSAGQGRDLQTLTIDSHDAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Db 901 DIDTLVANSAGQGRDLQTLTIDSHDAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Qy 961 HKTSGLOTLISLNSAMILNSRLVNSRRHTNHDISFAKRLQAKDOOKFASLESAAEVLYQF 1020
Db 961 HKTSGLOTLISLNSAMILNSRLVNSRRHTNHDISFAKRLQAKDOOKFASLESAAEVLYQF 1020
Qy 1021 APKYKPTNVWANAIGTISLNGSNASLYGTSGVDAYLINGOVEAIVGGFSGYSYSSFN 1080
Db 1021 APKYKPTNVWANAIGTISLNGSNASLYGTSGVDAYLINGOVEAIVGGFSGYSYSSFN 1080

Qy 1081 RANSLNSCANNNTNFGVYSRIFANQHEPDEFAQALGSDQSSLNFKSALLQDLNQSYHYLA 1140
Db 1081 RANSLNSCANNNTNFGVYSRIFANQHEPDEFAQALGSDQSSLNFKSALLQDLNQSYHYLA 1140
Qy 1141 YSAATRASYSYDFAFFRNALVKPSVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Db 1141 YSAATRASYSYDFAFFRNALVKPSVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Qy 1201 SANVEARYYYGDTSYFYMNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Db 1201 SANVEARYYYGDTSYFYMNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Qy 1261 LKLAKEVFLNLGVVYVYLNHLNLSNIGHFASNLGMRYSF 1296
Db 1261 LKLAKEVFLNLGVVYVYLNHLNLSNIGHFASNLGMRYSF 1296
RESULT 2
VAC2_HELPY
ID VAC2_HELPY STANDARD; PRT; 1287 AA.
AC Q48245;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49503 / 60190;
RX MEDLINE=94193753; PubMed=8144644;
RA Cover T.L., Tummuru M.K., Cao P., Thompson S.A., Blaser M.J.;
RT "Divergence of genetic sequences for the vacuolating cytotoxin among
Helicobacter pylori strains.";
RL J. Biol. Chem. 269:10566-10573(1994).
CC -1- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
ULCERATION AND GASTRIC LESIONS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U05676; AAA17657.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 1287 VACUOLATING CYTOTOXIN.
FT PROPEP ? 1287 POTENTIAL.
SQ SEQUENCE 1287 AA; 139041 MW; 0007370062FCB71F CRC64;
Query Match 94.5%; Score 6303.5; DB 1; Length 1287;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1227; Conservative 28; Mismatches 32; Indels 9; Gaps 2;
Qy 1 MEIOOTHRKINRPLVSLALVGVLSITPQOSHAAFFTTTIIPIAVGGTATGTAVGTVSGL 60
Db 1 MEIOOTHRKINRPLVSLALVGVLSITPQOSHAAFFTTTIIPIAVGGTATGTAVGTVSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLSLSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVRIQAGKGFNEFPNKEYDYLSLSLSSKIDGGWDMGNAAR 120
Qy 121 HYWKGGQONKLEVDMDKAVCTYTLISGLRNFTEGGDLVNMOKATLRLGOFNGNSFTSYKD 180
Db 121 HYWKGGQONKLEVDMDKAVCTYTLISGLRNFTEGGDLVNMOKATLRLGOFNGNSFTSYKD 180
Qy 181 SADRTTRVDFNAKNISIDNFEYINNRVSGSAGRKASSTVLTLOASEGITSKNAEISLYD 240
Db 181 SADRTTRVDFNAKNISIDNFEYINNRVSGSAGRKASSTVLTLOASEGITSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWVNGRLQVYVAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQA 300
Db 241 GATNLASSSVKLMGNVWVNGRLQVYVAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAAQA 300
Qy 301 GIANKTNICTLWOSAGLNIITAPPEGYKDKPNNTPSOSGAKNDKESAKNDKQESS 360
Db 301 GIANKTNICTLWOSAGLNIITAPPEGYKDKPNNTPSOSGAKNDKESAKNDKQESS 360
Qy 361 QNNSNTQVNPNSAQKTEVQTVIDGPPFAGGKDTVVNINRINTNADGTTIRVGGFKASL 420
Db 361 QNNSNTQVNPNSAQKTEVQTVIDGPPFAGGKDTVVNINRINTNADGTTIRVGGFKASL 420
Qy 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNQVGYALAGSSANFEF 480
Db 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNQVGYALAGSSANFEF 480
Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFTNDFSGVTDKVNINK 540
Db 481 KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTNGGFTNDFSGVTDKVNINK 540
Qy 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIQSQRINTVRLTETRSLFSGGV 600
Db 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSEDIQSQRINTVRLTETRSLFSGGV 600
Qy 601 KFKGGEKLVIDFYYSPNVDARINKEITNKLAFGPGSPWCTSKLMFNLLTGONA 660
Db 601 KFKGGEKLVIDFYYSPNVDARINKEITNKLAFGPGSPWCTSKLMFNLLTGONA 660
Qy 661 VMDYSQFSLTIQGDFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Db 661 VMDYSQFSLTIQGDFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Qy 721 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Qy 781 VRNTDDIKACMAIGDQSWVNPONKYLIKAKNKIGISTANGSKISVYVLGNSTPTE 840
Db 781 VRNTDDIKACMAIGDQSWVNPONKYLIKAKNKIGISTANGSKISVYVLGNSTPTE 840
Qy 841 NGGNTNLTPTTTSNARSANALAQNAFPAQSPATPNLVAIQHDFGIESVFELANRSK 900
Db 841 NGGNTNLTPTTTSNARSANALAQNAFPAQSPATPNLVAIQHDFGIESVFELANRSK 900
Qy 901 DIDTLVANSAGQGRDLQTLTIDSHDAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Db 901 DIDTLVANSAGQGRDLQTLTIDSHDAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Qy 961 HKTSGLOTLISLNSAMILNSRLVNSRRHTNHDISFAKRLQAKDOOKFASLESAAEVLYQF 1020
Db 961 HKTSGLOTLISLNSAMILNSRLVNSRRHTNHDISFAKRLQAKDOOKFASLESAAEVLYQF 1020
Qy 1021 APKYKPTNVWANAIGTISLNGSNASLYGTSGVDAYLINGOVEAIVGGFSGYSYSSFN 1080
Db 1021 APKYKPTNVWANAIGTISLNGSNASLYGTSGVDAYLINGOVEAIVGGFSGYSYSSFN 1080

Db 121 HWKGGQWNLKLEVDKDAVGYTKLSGRNFTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDENAKNISIDNEVEINNRVSGAGRKASSVTLTQASSEGITSKNAEISLYD 240
Db 181 SADRTTRVDENAKNISIDNEVEINNRVSGAGRKASSVTLTQASSEGITSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Qy 301 GIIANKKTNIGTDLWQASAGLNIIAPPEGYKDKPNNTPSQSGAKNDKESAKNDKQESS 360
Db 301 GIIANSKTHIGTDLWQASAGLNIIAPPEGYKDKPNNTPSQSG-----AKNDKQESS 352
Qy 361 QNNSNTQVINPNNSAKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Db 361 QNNSNTQVINPNNSAKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Qy 353 QNNSNTQVINPNNSAKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 412
Db 353 QNNSNTQVINPNNSAKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 412
Qy 421 TTNAHLHIGKGGVNLNOAGSRSLIVENLTGNITVDGRLRVNQVGYALAGSSANFEF 480
Db 421 TTNAHLHIGKGGVNLNOAGSRSLIVENLTGNITVDGRLRVNQVGYALAGSSANFEF 480
Qy 413 TTNAHLHIGKGGVNLNOAGSRSLIVENLTGNITVDGRLRVNQVGYALAGSSANFEF 472
Db 413 TTNAHLHIGKGGVNLNOAGSRSLIVENLTGNITVDGRLRVNQVGYALAGSSANFEF 472
Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDAAHTANFXGIDTNGGPFNTLDFSGVTDKVNINK 540
Db 481 KAGDTKNGTATFNNDISLGRFVNLKVDAAHTANFXGIDTNGGPFNTLDFSGVTDKVNINK 540
Qy 473 KAGVDTKNGTATFNNDISLGRFVNLKVDAAHTANFXGIDTNGGPFNTLDFSGVTDKVNINK 532
Db 473 KAGVDTKNGTATFNNDISLGRFVNLKVDAAHTANFXGIDTNGGPFNTLDFSGVTDKVNINK 532
Qy 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSDIGSQSRINTVRLTGTGRSLFSGGV 600
Db 541 LITASTNVAVKFNINELIVKNGISVGEYTHFSDIGSQSRINTVRLTGTGRSLFSGGV 600
Qy 533 LITASTNVAVKFNINELIVKNGISVGEYTHFSDIGSQSRINTVRLTGTGRSLFSGGV 592
Db 533 LITASTNVAVKFNINELIVKNGISVGEYTHFSDIGSQSRINTVRLTGTGRSLFSGGV 592
Qy 601 KFKGGEKLVIDEYFSPWNYEDARNIKVEITNKLAFAFGPQSGPWGTSKLMFNLLTGONA 660
Db 601 KFKGGEKLVIDEYFSPWNYEDARNIKVEITNKLAFAFGPQSGPWGTSKLMFNLLTGONA 660
Qy 593 KFKGGEKLVIDEYFSPWNYEDARNIKVEITNKLAFAFGPQSGPWGTSKLMFNLLTGONA 652
Db 593 KFKGGEKLVIDEYFSPWNYEDARNIKVEITNKLAFAFGPQSGPWGTSKLMFNLLTGONA 652
Qy 661 VMDYSQFNLTIQDGFINNQTINLYLVGGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
Db 661 VMDYSQFNLTIQDGFINNQTINLYLVGGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
Qy 653 VMDYSQFNLTIQDGFINNQTINLYLVGGKVATLSVGNAAAMFNNDIDSATGYKPLI 712
Db 653 VMDYSQFNLTIQDGFINNQTINLYLVGGKVATLSVGNAAAMFNNDIDSATGYKPLI 712
Qy 721 KINSAQDLIKKTEHVLKAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDCV 780
Db 721 KINSAQDLIKKTEHVLKAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDCV 780
Qy 713 KINSAQDLIKKTEHVLKAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDCV 772
Db 713 KINSAQDLIKKTEHVLKAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDCV 772
Qy 781 VRNTDDIKACGAIQDQSWNNPDNYKYLIKAWKNIGISKTAGSKISVYVLGNSTPTE 840
Db 781 VRNTDDIKACGAIQDQSWNNPDNYKYLIKAWKNIGISKTAGSKISVYVLGNSTPTE 840
Qy 773 VRNTDDIKACGAIQDQSWNNPDNYKYLIKAWKNIGISKTAGSKISVYVLGNSTPTE 832
Db 773 VRNTDDIKACGAIQDQSWNNPDNYKYLIKAWKNIGISKTAGSKISVYVLGNSTPTE 832
Qy 841 NGGNTNLPTNTTSNARSANALQNAFPAOPATPNLVAIQHDFGIESVFELANRSK 900
Db 841 NGGNTNLPTNTTSNARSANALQNAFPAOPATPNLVAIQHDFGIESVFELANRSK 900
Qy 833 NGGNTNLPTNTTNARFASALIKNAPFAH--SATPNLVAIQHDFGIESVFELANRSK 891
Db 833 NGGNTNLPTNTTNARFASALIKNAPFAH--SATPNLVAIQHDFGIESVFELANRSK 891
Qy 901 DIDTLYANGAGQGRDLQTLTLDSDHAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Db 901 DIDTLYANGAGQGRDLQTLTLDSDHAGYARKMIDATSAEITKOLNTATTTLNNIASLE 960
Qy 892 DIDTLYANGAGQGRDLQTLTLDSDHAGYARKMIDATSAEITKOLNTATTTLNNIASLE 951
Db 892 DIDTLYANGAGQGRDLQTLTLDSDHAGYARKMIDATSAEITKOLNTATTTLNNIASLE 951
Qy 961 HKTSGLOTLISLNSAMILNSRLVNLSSRRHTNHDSPAKLQALQKQFASLESAAEVLQF 1020
Db 961 HKTSGLOTLISLNSAMILNSRLVNLSSRRHTNHDSPAKLQALQKQFASLESAAEVLQF 1020
Qy 952 HKTSGLOTLISLNSAMILNSRLVNLSSRRHTNHDSPAKLQALQKQFASLESAAEVLQF 1011
Db 952 HKTSGLOTLISLNSAMILNSRLVNLSSRRHTNHDSPAKLQALQKQFASLESAAEVLQF 1011
Qy 1021 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGFGSGYSSFN 1080
Db 1021 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGFGSGYSSFN 1080
Qy 1012 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGFGSGYSSFN 1071
Db 1012 APKYEKPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGFGSGYSSFN 1071
Qy 1081 RANSLNSGANTNFVYKRIFANOHEPDEAQAQALGDSQSSLNFKSALLQDLNQSYHYLA 1140
Db 1081 RANSLNSGANTNFVYKRIFANOHEPDEAQAQALGDSQSSLNFKSALLQDLNQSYHYLA 1140
Qy 1072 QANSLNSGANTNFVYKRIFANOHEPDEAQAQALGDSQSSLNFKSALLQDLNQSYHYLA 1131
Db 1072 QANSLNSGANTNFVYKRIFANOHEPDEAQAQALGDSQSSLNFKSALLQDLNQSYHYLA 1131
Qy 1141 YSAATRASGYDFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Db 1141 YSAATRASGYDFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
Qy 1132 YSAATRASGYDFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1191
Db 1132 YSAATRASGYDFAFERNALVKPSGVSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1191
Qy 1201 SANVEARYYYGDTSYFYMNAGVLOQFAHVGSNNAAASLNTFKVNAARNPLNTHARVMGGE 1260
Db 1201 SANVEARYYYGDTSYFYMNAGVLOQFAHVGSNNAAASLNTFKVNAARNPLNTHARVMGGE 1260

Db 1192 SANVEARYYYGDTSYFYMNAGVLOQFAHFANFGSSNAVSLNTFKVNATRNPLNTHARVMGGE 1251
Qy 1261 LKLAKEVFLNLGVVYVHLNLIISNIGHFASNLGMRYSF 1296
Db 1252 LKLAKEVFLNLGVVYVHLNLIISNIGHFASNLGMRYSF 1287
RESULT 3
VAC4_HELPY
ID VAC4_HELPY STANDARD; PRT: 1291 AA.
AC Q48258;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=185-44;
RX MEDLINE=94333650; PubMed=8057855;
RA Haas R., Schmitt W.;
RT "Genetic analysis of the Helicobacter pylori vacuolating cytotoxin:
RT structural similarities with the IgA protease type of exported
RT protein.";
RL Mol. Microbiol. 12:307-319(1994).
CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: 226883; CAA81528.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 33
FT CHAIN 34 ?
FT PROPEP ? 1291
SQ SEQUENCE 1291 AA; 139635 MW; ECAS6A61CAE36669 CRC64;

Query Match 93.3%; Score 6222.5; DB 1; Length 1291;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1202; Conservative 51; Mismatches 38; Indels 5; Gaps 1;
Qy 1 MEIQOTHRKINRPLVSLALVGVLSITPOQSHAAFFTVIIIPAIYGGIATGAVGTVSGL 60
Db 1 MEIQOTHRKINRPLVSLALVGVLSITPOQSHAAFFTVIIIPAIYGGIATGAVGTVSGL 60
Qy 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEDLYKSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEDLYKSLSSKIDGGWDMGNAAR 120
Qy 121 HWVVGKGOONKLEVDKDAVGYTKLSGRNFTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
Db 121 HWVVGKGOONKLEVDKDAVGYTKLSGRNFTGGDLVNMOKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDENAKNISIDNEVEINNRVSGAGRKASSVTLTQASSEGITSKNAEISLYD 240
Db 181 SADRTTRVDENAKNISIDNEVEINNRVSGAGRKASSVTLTQASSEGITSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATNLASSSVKLMGNVWGRLOQYVAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300

QY 301 GIANKKTIGTLDLWQSLNIIAPPEGGYKDKPNTPSQGAKNDKNEAKNDKQESS 360
DB 301 GIANSKTHIGTLDLWQSLNIIAPPEGGYKDKPKDPS-----NTTONNANNQONSA 355
QY 361 QNNSNTQVNPNSAQKTEVQPTQVIDGPFAGKDTVNVNINRINTNADGTIRVGGFKASL 420
DB 356 QNNSNTQVNPNSAQKTEIQPTQVIDGPFAGKDTVNVNINRINTNADGTIRVGGFKASL 415
QY 421 TTNAAHLHIGKGVNLSNOAGRSILVENLTGNTITVDGPLRVNNOVGGYALAGSSANPEF 480
DB 416 TTNAAHLHIGKGVNLSNOAGRSILVENLTGNTITVDGPLRVNNOVGGYALAGSSANPEF 475
QY 481 KAGDTKNGTATFNNDISLGRFVNKLVDAHTANFKGIDTNGGFFNTLDFSGYTDKVNINK 540
DB 476 KAGDTKNGTATFNNDISLGRFVNKLVDAHTANFKGIDTNGGFFNTLDFSGYTDKVNINK 535
QY 541 LITASTNAVAKNFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSLFSGGV 600
DB 536 LITASTNAVAKNFNINELIVKNGISVGEYTHFSEIDIGSQSRINTVRLTGTFRSLFSGGV 595
601 KFKGGEKLVDFEYSPWNYFDARNIKNVEITNKLAFGPQGGPWGTSKLMFNNTLIGQNA 660
596 KFKGGEKLVDFEYSPWNYFDARNIKNVEITRKFASTPENPWGTSKLMFNNTLIGQNA 655
QY 661 VMDYSQFSLTIQGDFFINNQGTTINLVRGKGVATLSVGNAAAMFNNDIDSATGYKPLI 720
DB 656 VMDYSQFSLTIQGDFFINNQGTTINLVRGKGVATLSVGNAAAMFNNDIDSATGYKPLI 715
QY 721 KINSQDLKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEDEQFKERLALYNNNRMDTCV 780
DB 716 KINSQDLKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEDEQFKERLALYNNNRMDICV 775
QY 781 VNTDDIKAGMAIGDQSVNPNPNKYLLIGKAWKNIGTSKTANGSKISVYVLGNSTPTE 840
DB 776 VNTDDIKAGTAIGQSVNPNPNKYLLIGKAWKNIGTSKTANGSKISVYVLGNSTPTE 835
QY 841 NGGNTTNLTNTTSNARSANNAQAAPPQATPNLVAIQHDFGTIESVFELANRSK 900
DB 836 KGGNTTNLTNTTSNARSANNAQAAPPQATPNLVAIQHDFGTIESVFELANRSK 895
QY 901 DITDLYANSAGORLLQTLTLDSDHAGYARKMIDATSAEITKQNTATTTLNINIASLE 960
DB 896 DITDLYANSAGORLLQTLTLDSDHAGYARQMDTSTGEITKQNLNAAATTTLNINIASLE 955
QY 961 HKTSGLOTLSLNNMLNRLNRRHTNHIDSPAKRLQALKDQKFALESAAEVLQF 1020
DB 956 HKTSSLOTLSLNNMLNRLNRRHTNHIDSPAKRLQALKDQKFALESAAEVLQF 1015
QY 1021 APKYEPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFPNN 1080
1016 APKYEPTNVWANAIGGTSNLNGSNASLYGTSAGVDAYLNGQVEAIVGGFGSYGYSFPNN 1075
QY 1081 RANSLNGANNNTFGYYSRIFANQHEFDEFAQALGSDQSSLNFKSALLQDLNQSYHYLA 1140
DB 1076 QANSLNGANNNTFGYYSRIFANQHEFDEFAQALGSDQSSLNFKSALLQDLNQSYHYLA 1135
QY 1141 YSAATRASGYDFAFRNALVLKPSGVSYNHLGSNFKSNSTNOVALKNGSSQHLFNA 1200
DB 1136 YSAATRASGYDFAFRNALVLKPSGVSYNHLGSNFKSNSTNOVALKNGSSQHLFNA 1194
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFARHYGSSNAAASLNTFKYNAARNPLNTHARVNMGGE 1260
DB 1195 SANVEARYYYGDTSYFYMNAGVLQEFARHYGSSNAAASLNTFKYNAARNPLNTHARVNMGGE 1254
QY 1261 LKLAKEVFLNLGVYVHLNLSINIGHFASNLGMRYSF 1296
DB 1255 LKLAKEVFLNLGVYVHLNLSINIGHFASNLGMRYSF 1290

RESULT 5
VACA_HELPJ
ID VACA_HELPJ STANDARD; PRT; 1288 AA.

Q92KW5;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA OR JHP0819.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001511; AAD06400.1; -;
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal; Complete proteome.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 ? VACUOLATING CYTOTOXIN.
FT PROPEP ? 1288 POTENTIAL.
SQ SEQUENCE 1288 AA; 139131 MW; 2448159DFC5F32B9 CRC64;

Query Match 91.6%; Score 6108; DB 1; Length 1288;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1181; Conservative 53; Mismatches 54; Indels 8; Gaps 2;

QY 1 MEIQOTHRINRPLVSLALVGVSTTPQOSHAAFTTIIIPAVGGIATGTAVGTVSG 60
DB 1 MEIQOTHRINRPLVSLVLAGALISAIPQESHAAFTTIIIPAVGGIATGTAVGTVSG 60
QY 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDGNAAR 120
DB 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDGNAAR 120
QY 121 HWYKGGQGNKLEVDKMDAVGTYTSLGLNFTGGDLVDVNNQKATRLGQFNNGNSFTSYKD 180
DB 121 HWYKGGQGNKLEVDKMDAVGTYKLSGLNFTGGDLVDVNNQKATRLGQFNNGNSFTSYKD 180
QY 181 SADRTTRVDFNAKNSIDNFEVNNRVGSGAGRKASSTVLTTLQASEGITSSKNAEISLYD 240
DB 181 SADRTTRVDFNAKNSIDNFEVNNRVGSGAGRKASSTVLTTLQASEGITSSKNAEISLYD 240
QY 241 GATNLASSSVKLMGVNWRLOQYVAYLAPSYSTINTSKYTCGEVNFHLYTGDCKNAAQA 300
DB 241 GATNLASSSVKLMGVNWRLOQYVAYLAPSYSTINTSKYTCGEVNFHLYTGDCKNAAQA 300
QY 301 GIANKKTIGTLDLWQSLNIIAPPEGGYKDKPNTPSQGAKNDKNEAKNDKQESS 360
DB 301 GIANSKTHIGTLDLWQSLNIIAPPEGGYKDKPNTPSQGAKNDKNEAKNDKQESS 354
QY 361 QNNSNTQVNPNSAQKTEVQPTQVIDGPFAGKDTVNVNINRINTNADGTIRVGGFKASL 420
DB 361 QNNSNTQVNPNSAQKTEVQPTQVIDGPFAGKDTVNVNINRINTNADGTIRVGGFKASL 420

Db 355 - NNSNTEVINPPNNTQKTETPTQVIDGPFAGGKDTVVNIFHLNKTADGTIKVGGFKASL 413
 Qy 421 TTNAHLHIGKGGVNLNOAGSRSLIVNLGNITVDGPLRVNNOVGAYALAGSSANPEF 480
 Db 414 TTNAHLNIGKGGVNLNOAGSRLLVNLGNITVDGPLRVNNOVGAYALAGSSANPEF 473
 Qy 481 KAGDTKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNCGFNTLDFSGVTDKVNINK 540
 Db 474 KAGVDTKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTGNCGFNTLDFSGVTDKVNINK 533
 Qy 541 LITASTNVAVKNFNLINELIVTNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV 600
 Db 534 LITASTNVAVKNFNLINELIVTNGISVGEYTHFSEDIGSQSRINTVRLTGTSLFSGGV 593
 Qy 601 KFKGGEKLVIDEFYSPNYPNADRNKVEITNKLAFQPGQSPWCTSKLMFNNTLGLGNA 660
 Db 594 KFKGGEKLVINDFYSPNYPNADRNKVEITRFASTPENPGCTSKLMFNNTLGLGNA 653
 Qy 661 VMDYSQFSLTIQDGFINNCGTINYLVRGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
 Db 654 VMDYSQFSLTIQDGFINNCGTINYLVRGKVATLSVGNAAAMFNNDIDSATGYKPLI 713
 Qy 721 KINSADLIKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 780
 Db 714 KINSADLIKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNRMDTCV 773
 Qy 781 VRNTDDIKACGMAIGDQSMVNNPNYKYLIGKAWNIGISKTAGSKISVYVLGNSTPTE 840
 Db 774 VRNTDDIKACGMAIGDQSMVNNPNYKYLIGKAWNIGISKTAGSKISVYVLGNSTPTE 833
 Qy 841 NGGNTNLTPTNTNARSANALQAPPAQPSATPNLVAIQHDFGTIESVFELANRSK 900
 Db 834 NGGNTNLTPTNTNARSANALQAPPAQPSATPNLVAIQHDFGTIESVFELANRSK 892
 Qy 901 DIDTLYNSGAGQDRDLTLTLDSDHAGYARKMIDATSAEITKOLNTATTTLNNAISLE 960
 Db 893 DIDTLYNSGAGQDRDLTLTLDSDHAGYARKMIDTSTGEITKOLNATDALNNVASLE 952
 Qy 961 HKTSGLOTLSLSNAMILNLSRLVNLRRTHNIDSFARKLQALQKQKFALESAAEVLQF 1020
 Db 953 HKQSGLOTLSLSNAMILNLSRLVNLRRTHNINSAFQRLQALQKQKFALESAAEVLQF 1012
 Qy 1021 APKYEKPTNVAANAIGTSLNNGSNASLYGTSGVAYLNGQVETIAVGGFSGYSYSSFN 1080
 Db 1013 APKYEKPTNVAANAIGTSLNNGSNASLYGTSGVAYLNGQVETIAVGGFSGYSYSSFN 1072
 Qy 1081 RANLSNCGANTNFGVYSRIFANQHEFDEAQAAGLQKQKFALESAAEVLQF 1140
 Db 1073 QANLSNCGANTNFGVYSRIFANQHEFDEAQAAGLQKQKFALESAAEVLQF 1132
 Qy 1141 YSAATRASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1200
 Db 1133 YSAATRASGYDFAFFRNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSSOHLFNA 1192
 Qy 1201 SANVEARYYIGDTSYFYNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1260
 Db 1193 SANVEARYYIGDTSYFYNAGVLOEFAHVGSNNAASLNTFKVNAARNPLNTHARVMGGE 1252
 Qy 1261 LKLAKEVFLNGLVYVHLNLSNIGHFASNLGMWYSF 1296
 Db 1253 LQLAKEVFLNGLVYVHLNLSNIGHFASNLGMWYSF 1288
 RESULT 6
 VAC3_HELPY
 ID VAC3_HELPY STANDARD; PRT; 1310 AA.
 AC Q48253;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Vacuolating cytotoxin precursor.
 GN VACA.
 OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TX30A;
 RX MEDLINE=9535366; PubMed=7629077;
 RA Atherton J.C., Cao P., Peek R.M. Jr., Tummuru M.K., Blaser M.J.,
 RA Cover T.L.;
 RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
 RT Association of specific vac types with cytotoxin production and
 RT peptic ulceration.";
 RL J. Biol. Chem. 270:17771-17777(1995).
 CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
 CC ULCERATION AND GASTRIC LESIONS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U29401; AAA86834.1; -
 DR InterPro: IPR003842; Vaca.
 DR Pfam: PF02691; Vaca: 1.
 KW Cytotoxin; Toxin; Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 ? VACUOLATING CYTOTOXIN.
 FT PROPEP ? 1310 POTENTIAL.
 SQ SEQUENCE 1310 AA; 141988 MW; 1BC21FE3D435F981 CRC64;

 Query Match 78.58; Score 5237; DB 1; Length 1310;
 Best Local Similarity 77.98; Pred. No. 6.2e-270;
 Matches 1045; Conservative 88; Mismatches 131; Indels 78; Gaps 13;
 Qy 1 MEIOOTHRKINRPLVSLALVGLVSI-----TP-----QQSHAAFTVTIIIPAVGGIATG 51
 Db 1 MEIOOTHRKINRPLVSLALVGLVGLMTELGANTPNPHSESRATFTVTIIIPAVGGIATG 60
 Qy 52 TAVGTVSGLLSWGLQKQAEANKTPDKVWRIQAGKGFNEFPNKEYDLYSLSSKIDG 111
 Db 61 AAVGTVSGLLSWGLQKQAEANKTPDKVWRIQAGKGFNEFPNKEYDLYSLSSKIDG 120
 Qy 112 GWDHGNARHYVWVGQONKLEVDMDKAVGTYTLGLRNFTGGDLVNMOKATLRLGQFN 171
 Db 121 GWDHGNARHYVWVGQONKLEVDMDKAVGTYTLGLRNFTGGDLVNMOKATLRLGQFN 180
 Qy 172 GNSFTSYKDSADRTVRVDFNAKNSIDNFVEINNRVSGAGRKASSTVLTQLQASEGITS 231
 Db 181 GNSFTSYKDSADRTVRVDFNAKNSIDNFVEINNRVSGAGRKASSTVLTQLQASEGITS 240
 Qy 232 KNAEISLYDGTATLNLAS---SVKLMGNVWGRLOYVGYLAPSYSTINTSKVTGEVNFN 288
 Db 241 ENAEISLYDGTATLNLSSNSQSDVLYGKVMGRLOYVGYLAPSYSTINTSKVTGEVNFN 300
 Qy 289 HLTVDGKNAAGIITANKKTNIPTGLDQWAGLNTIAPPEGYKDKPNNTPSQSGAKNDK 348
 Db 301 HLTVDGKNAAGIITANKKTNIPTGLDQWAGLNTIAPPEGYKDKPNNTPSQSGAKNDK 351
 Qy 349 NESAKNDQESSNNSTQVFNPPNSAOKTEVQPTQVIDGPFAGGKDTVVNINRINTNAD 408
 Db 352 -----QN-----NPKNDAGKTEIQPTQVIDGPFAGGKDTVVNIFHLNTRAD 392
 Qy 409 GTIRVGGFKASLTNAHLHIGKGGVNLNOAGSRSLIVNLGNITVDGPLRVNNOVG 468
 Db 393 GTIRVGGFKASLTNAHLHIGKGGVNLNOAGSRSLIVNLGNITVDGPLRVNNOVG 452
 Qy 469 YALAGSSANPEFAKDTDKNCTATFNNDISLGRFVNLKVDHAHTANFKG----- 516
 Db 452 YALAGSSANPEFAKDTDKNCTATFNNDISLGRFVNLKVDHAHTANFKG----- 516

Db 453 AAIAGSSANFEKAGEDTNNATATFNNDIHLGKAVNLRVDAHTANFNNGNIYLGKSTNLRV 512
QY 517 -----IDT--GNGFN--TLDFSGVTDKVNINKLITASTNVAVKNFNINELIVKTN 563
Db 513 NGHTAHFKNIDATKSDNGLNTSTLDFSGVTDKVNINKLITATNVAVKNFIDIKELVVITR 572
QY 564 GTSVGEYTHFSDSQSQRINVRLETGRSFGVKGKGEKVIDEFYSPMNYFDA 623
Db 573 VQSGGYTFIFGNIKSGRIGVVSQTSYSPAYSGVTEKGGKLVDEIYHAPNYPDA 632
QY 624 RAIKWEITNKLAFGPQSPWGTSKLMFNLLIGQNAVMDYSQFSLNLTQGFNNOGTI 683
Db 633 RNVTDVEINKLTFAGPAGIAGTKGLMFNLLTNSNASMDYDKDLDLTQGHFTNQGTM 692
QY 684 NYLVGGKVATLSVGNAAAMFNDDISATGYKPLIKINSQDLIKNTEHVLAKAKIIG 743
Db 693 NLFVODGRVATLNAGHQASMIENLVDSITGYKPLIKINNAQNLTKNKEHVLVARNID 752
QY 744 YGNVSTGTNG-----ISNVLEEQFKERLALYNNNNRMDTCVVR--NTDDIKACGMAI 794
Db 753 YNLV--GVQASDYNTISASNTNLQEQFKERLALYNNNNRMDTCVVRKONLNDIKACGMAI 810
795 GQSVWNPNDYKYLIGKAWKIGISKANGSKISVIYLCNSTPTENGNTTNLTNTTS 854
Db 811 GQSVWNPNDYKYLIGKAWKIGISKANGSKISVIYLCNSTPTENGNTTNLTNTTN 870
QY 855 NARSANALQANPAQPSATNLVAINOHDFGTIESVPELANRSKDIDTLVANSQAQR 914
Db 871 NARFASYALIKNAPAH--SATNLVAINOHDFGTIESVPELANRSKDIDTLVANSQAQR 929
QY 915 DLLQTLIDSDHAGYARKMIDATSAEITKQLTNTATTTLNNTIASLEHKTSGLTLSLSNA 974
Db 930 DLLQTLIDSDHAGYARTMIDATSAEITKQLTNTATTTLNNTIASLEHKTSGLTLSLSNA 989
QY 975 MTLNRLNLSRHHNHIDSFARLQALQDKQFASLESAAEVLVYQFAPKYKPTNWNANA 1034
Db 990 MTLNRLNLSRHHNHIDSFARLQALQDKQFASLESAAEVLVYQFAPKYKPTNWNANA 1049
QY 1035 IGGTSLNNGSNASLGTSGAGVDAYLNGQVEAIVGGFSGYSGSFNNRANSLNSGANNTNF 1094
Db 1050 IGGTSLNNGSNASLGTSGAGVDAYLNGQVEAIVGGFSGYSGSFNSQANSLNSGANNTNF 1109
QY 1095 GYVSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNOSYHYLAYSATTRASGYDFA 1154
Db 1110 GYVSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNOSYHYLAYSATTRASGYDFA 1169
QY 1155 FFRNALVLKPSGVSYNHLGTFNFKSNSTNOVALKNGSSQHLFNASANVEARYYTGDS 1214
Db 1170 FFRNALVLKPSGVSYNHLGTFNFKSNSTNOVALKNGSSQHLFNANANVEARYYTGDS 1228
1215 YFYMAGVLQEPFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGELKLAKFVNLGVV 1274
Db 1229 YFYMAGVLQEPFAHVGSNNAASLNTFKVNAARNPLNTHARVMMGELKLAKFVNLGVV 1288
QY 1275 YLHNLSITNGHFAASNLGMRYSF 1296
Db 1289 YLHNLSITNGHFAASNLGMRYSF 1310
RESULT 7
OMP_RICRI STANDARD: PRT: 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

Rickettsiaceae; Rickettsieae; Rickettsia.
NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
rickettsii is encoded by an unusually long open reading frame:
evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC !- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC !- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
LAYER WITH HEXAGONAL SYMMETRY.
CC !- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
DR EMBL; X16353; CAA34403.1; --
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
SQ
Query Match 5.0%; Score 335.5; DB 1; Length 1654;
Best Local Similarity 20.5%; Pred. No. 4.2e-10;
Matches 292; Conservative 169; Mismatches 516; Indels 445; Gaps 64;
QY 138 DAVGTYTSLGLRNFRTGGDLDDV-NMQKATLRL-----GQ---FNGN----- 173
Db 203 NAVIANGVNGTSLNVTNGFTQVSNKSPATVKAINIADQGGIIFTDANNANTLNQAGGIT 262
QY 174 -SFTSYKDSADTRTRVDFNAKNISIDNFVEINNRVGS-GAGRKASSTVTLTQASEGITS 231
Db 263 INF-----GTDTGTRLVLLSKHAAATNF---NITSLGLNKLGVTEFTVAVDGLTAN 314
QY 232 KNAEISLDGATLNLAASSSVKLMGNVMMGRLOYVGAYLAPSYSTINTSKVTGVNNEHL- 290
Db 315 AGAANAVI--GTNNGAGRAAGFVSVNDGKVTATIDQVYAKDMVQSANATGVNFRHIV 372
QY 291 -----TVGDKNAQAAGIANKKTNIGTLDMQSAAGLNIIAPP-----EGGYKDKPNTPS 340
Db 373 DVGADGTTAFKTA--SKVTITQDSNFGNTDFGNLAA-QIKVPNAITLTGFTGDSNPGN 430
QY 341 QSGA-----KNDKNESAKNDKQESSQNN----- 363
Db 431 TAGVITFDANGTLESASADANVAVTNNITAEASGAGVVLQSGTHAAELRLNAGSIFKL 490
QY 364 SNTQVNPNSAQKTEVQPTVIDGPFAGKGTVVNININTNADGTIRVGG----- 415
Db 491 ADGTVIN-----GKVNQTLVGGALAAAGTITL-----DGSATITDIGNAGGA 533
QY 416 -----FKASLTNAAHLHGKGVNLSNQASGRSLIVENLTGNTITVD----- 457

[illegible]

RESULT 8

FT	VARIANT	1139	1139	R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT	CONFLICT	353	354	KD -> GH (IN REF. 3).
FT	CONFLICT	776	776	F -> S (IN REF. 3).
FT	CONFLICT	1159	1159	E -> D (IN REF. 3).
FT	CONFLICT	1177	1177	G -> S (IN REF. 3).
FT	CONFLICT	1492	1492	H -> R (IN REF. 3).
SQ	SEQUENCE	1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;		
Query Match				
Best Local Similarity 5.0%; Score 335; DB 1; Length 1655;				
Matches 274; Conservative 177; Mismatches 531; Indels 420; Gaps 59;				
Qy	146	SCLRNFTGGDLVNNMOK	ATRLGQFNGSFTSYKDSADRTTRVDNFNAKNISIDNEFEIN	204
Db	211	NGTLNVTNGFIKVSXSFATVNVINIGDGGIMFNTDADNVNTLNQANGATI	----	TFN 266
Qy	205	NRVSGS	-----AGRKASSTVLTQAS	-----EGI
Db	267	GTDTGRLVLLSKNAATDFNVTSGLGNLKGII	EFNTVAVNGQLKANAGANA	AVIGTNN 326
	246	LASSSVKLMGNVMGRLVGLAYLAPSYSTINTSKVTGEVNFHL	-----	TVGDKNAA 298
	327	GAGRAAGFVWSVNGKATIDQVYAKDKVIO	SANAVGOVNRHIVDVGTGDTTAFKTA	386
Qy	299	QAGIIANKKTNIGTLDLWQAGLNIAPPE	-----GGYKDKPNNTP	-----SOS 342
Db	387	SKVAITQNSFGTDFGNLAA	-----QIIIVPNTMTLNGNFTGDASNPGNTAGVITFDANGTIA	444
Qy	343	GAKNDKESAKNDQESSONSNTOVINPPNSAQ	-----	KTEVQPTQ 384
Db	445	SASADANAVTNNITAI	ESAGVGVQLSCTHAELRLGNAGSVFKLADGTGVINGKVNQTA	504
Qy	385	VLDGPFAGGKDFV	-----VNINRINTNADGTIRVGGF	-----KASLTNAAHLHICGGVN 435
Db	505	LVGGALAAAGTITLDGASATITDIGNAGGAALQGTITLANDATKTLTLCGANI	-----IGANGT 563	
Qy	436	LSNQASGRSLIVENITGNITDGLRVNNQVGYALAGSSANFEF	-----KAGDTKNG	489
Db	564	INFOANGGTIKLTSQNNIVDFDLAIATDQTVGVDDASLTAQTLTNGKIGTVGANNK	623	
Qy	490	-----TATNNDISLGRFV	-----	NLKVD 508
Db	624	TLGQFNIGSSKTVLSDGDVA	INELVIGNNAVQFAHNTVILTRTTNAAGQGGKIIFNPVVN	683
Qy	509	AHTANFKGIDTNG	-----GFNTLDSGVTDKVINKILITA	STNVAVKFNINELIV 560
Db	684	NNTTATGTLNLSATNPLAEINFGSKGAANVDTVLNVGKGVNLYATNITTTDANVGSFIF	743	
Qy	561	KTNGISVGEYTHFSEDIGSQ	-----SRINTVRLETGTRSLFSGGVKFKGGEKLVIDEFYFSPW	618
	744	NAGGTNI	-----VSGTVGGQQGNKENTVALDNGTTVKFLGNATFNQNTTIAANSTLIQGG	798
Qy	619	NY	-----FDRNKINVEITN	-----KIAFGQPS
Db	799	NYTADFVASADGTGI	-----VEFVNTGPITVTLNKAAPVNAKQITVSGPGNVVINEIGNAG	856
Qy	659	N	-----AVMDYSQFSLTIQGFENNQGTINYLVRGKGVATFSLVSGAAAMFNNDI	-----DSAT 713
Db	857	NYHGAVTDTIAFENSL	-----GAVVFLPRG	-----IPENDAGNRIPLIKTGVGNKAT 906
Qy	714	GYFKPLIKINSADQILKNTHEVLLKAKIYG	-----NVSTGTNGISNVNLEE	-----761
Db	907	GFDVPSVILGVDSVIADQGVIGDQNNIVGLGSDNDIIVNATLTLYAGIGTINNQGTV	966	
Qy	762	-----	-----QKERIAL	-----YNN
Db	967	TLSGGIPNTPGTVYGLGTIGIGASKEFKQVFTTDDYNNLGNIIATNATINDGVVTTGGIAG	1026	
Qy	790	CG	-----MAIGDQSMVNNPDNYK	-----LTGKWKNGIKSTANGSKISVYVLGNS
Db	1027	IGFDGKITLGS	-----VNGNGNRFVDCILSHSTSMIGTTKANNG	-----TVTYLGNAFVGNIG 1080

QY	837	---TPTEN	-----GGNTT	-----NLPT 850
Db	1081	DSDTPVASVRFSGDGGAGLOGNIYSQVDFGTYNLIGSINNVILGGGTAINGKINLRT	1140	1140
QY	851	NNTSNARSANALAQNAQPAQSPATPNLVAIQHDEGTTIESPELANRS	-----899	899
Db	1141	NLTFTA-SGTSTWGNNT	-----SIETTLTLANGNIGNIVILEGAQ	1179
QY	900	-----KDIDTLYAN-SGAOGRDLQ	-----TLIDSHDAGYARKMD	935
Db	1180	VNATTTGTTIKVQDNANANFSGTQTYTLIQGARPNGTLGGPNFVVTGSNRVNVGLIR	1239	1239
QY	936	ATYSAEITKQLNTA	-----TTLNIAISLEHKTSGLOTLSL	972
Db	1240	AAQDYVITRTNAENVVNDIANSSFGCAGPGVQGVNTTFVNATNTAAYNNLLAKNSAN	1299	1299
QY	973	NAMILNSRLVNLSSRHTNHIDSFAKRLQALKQKQFASL	-----1010	1010
Db	1300	SANFGAIVTDTSAAITNAQLDVAKDIAQLGNRLGALRYLGTPTAEMAGPEAGAIPAA	1359	1359
QY	1011	-----ESAAEVLVQFAPKYEKP	-----TNVWANAIGGTSNLNNGSNASLYGTSAGVDAYLNGOV	1063
Db	1360	VAGDEAVDNVAYGI	-----WAKPFYTDHQSKKGLA	-----GYKAKTTGVVIGLDTLANDNL 1413
QY	1064	EAIVGGFGSYGYSNNRANSLSNG	-----ANNTNFGVYSRIFANQHEFDEAOG	-----ALGSD 1118
Db	1414	--MIG--AAIGITKTDIKHQDYKKGDKTDVNGFSFSLYG	---AQQLVKNFFAQSIAFSL	1466
QY	1119	QSSLNFKSALLQDLQSY	-----HYLAYSATRASGYDFAFFRNALVLPKSVGVSYN	1171
Db	1467	NOVKKRSQRYFFDANGNMSKQIAAGHYDNMTFEGNLTGVGDYNAMQGVLY	-----TPMAGLSYL	1525
QY	1172	HLGSTNFKSNSTNOVAKNGSS	1193	
Db	1526	KSSDENYKETGTT	VANKQVNS	1546
RESULT 9				
OMP	RICJA	STANDARD;	PRT; 1656 AA.	
ID	OMP	RICJA		
AC	00653			
DT	30-MAY-2000	(Rel. 39, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Outer membrane protein B precursor	(168 kDa surface-layer protein)		
DE	(Surface protein antigen)	(Cell surface antigen 5) (Sca5) (rOmpB)		
DE	(rOmp B)	(Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).		
GN	OMP			
OS	Rickettsia japonica			
OC	Bacteria; proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Rickettsiae; Rickettsia			
OX	NCBI_TaxID=35790;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=YH;			
RA	Uchiyama T.;			
RT	"Sequencing of the gene encoding the protein romp B of Rickettsia japonica."			
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIOLENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY SIMILARITY).			
CC	-!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.			
CC	-!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on			


```

Db 1281 TGTAVANN-----EGTWNLGSSEGN--TGMLEVVNNS-----AFN--NRGEFLDND-- 13210
QY 1157 RNALVLKPSGVSYNHLSSTNFKSNSTNOVALKNGSSSQHLFNASAVEARYYY----- 12110
Db 1326 KNAVHINQS--GTLYN--TGHMNTISNSHGAVNMGGNGRFTNDGTIDVSAKSLVVSANNA 13813
QY 1211 GDTSYFYMNA--GVQLQEFPAH-----VGSNNAASL 12317
Db 1384 GQONAEFFNQDNGVI--NFDHDSASAVKVYTHSNFIAQNDGIMNISGTGAVAMEGDKNAQLV 14412
QY 1238 NTFKVAARNPLNTHARVNMGGEL---KLAREVELNLGW-----YLHNLISNIGHEA 12817
Db 1443 NNGCTIML--GTAGTTDTGMIGMOLDANATADAVIENNGTINIFANDSFAFSLVGTGVGHV 15010
QY 1288 SN 1289
Db 1501 NN 1502

RESULT 11
ID OMPA_RICRI STANDARD; PRT; 2249 AA.
AC P15921:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (rOmpA) (rOmp A).
DE GN OMPA.
DE OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR;
RX MEDLINE=90354033; PubMed=2117568;
RT Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences.";
RT Infect. Immun. 58:2760-2769(1990).
CC -! FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -! SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC -! LAYER WITH HEXAGONAL SYMMETRY.
CC -! PTM: GLYCOSYLATED (PROBABLE).
CC -! SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collabora-
tion between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on
the use by non-profit institutions as long as its content is in no
way modified and this statement is not removed. Usage by and for commer-
cial entities requires a license agreement (See http://www.isb-sib.ch/announ-
ce or send an email to license@isb-sib.ch).

-----
Cc EMBL: M31227; AAA26380.1; -.
Cc PIR: A41477; A41477.
Cc InterPro: IPR003858; rOmpA_rOmpB.
Cc Pfam: PF02708; rOmpA_rOmpB; 1.
Cc Kw Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.
Cc FT SIGNAL 1 28 POTENTIAL.
Cc FT CHAIN 29 2249 OUTER MEMBRANE PROTEIN A.
Cc FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
Cc FT REPEAT 212 286 A (TYPE I).
Cc FT REPEAT 287 358 B (TYPE II).
Cc FT REPEAT 359 430 C (TYPE II).
Cc FT REPEAT 431 505 D (TYPE I).
Cc FT REPEAT 506 577 E (TYPE II).
Cc FT REPEAT 578 652 F (TYPE I).
Cc FT REPEAT 653 724 G (TYPE II).
Cc FT REPEAT 725 799 H (TYPE I).
Cc FT REPEAT 800 874 I (TYPE I).
Cc FT REPEAT 875 949 J (TYPE I).

```

Query Match 4.7%; Score 312.5; DB 1; Length 1300;
Best Local Similarity 21.1%; Pred. No. 4.9e-09;
Matches 275; Conservative 160; Mismatches 491; Indels 379; Gaps 61;

Db 44 GNTDFGNLA-----AQIKVPNAI---TLTG--NFTG---DASNPGNTAGVITF 83
Qy 171 NGNSFTSYKDSADRTVRDFNAKNIISIDNFEINNRVSGAGRKGASSTVLTLQASEGITS 230
Db 84 DANG-TLESASAD-----ANVATNITATEASAGVVLQSGT-----HAEELRLG 128
Qy 231 DKNAEISLYDGATLANLASSVYKMGNYMRLQVYGLAPSYSTINTSKVTGEVFNHL 290
Db 129 NAGSIFKLADGTVINGKVNQALVG-----GALAAAGTITLDGSAITIGDIG--- 174
Qy 291 TVGDKNAQAQIANKNTNIGTLDLWOSAGINIIAPPEGYKDKPNPTPQSGAKNDKNE 350
Db 175 NAGGAALQRITTLANDAKTLTL-----GGANII-----GAGGCTIDLQA 214
Qy 351 SAKNDKQESSONSNTQVINPPNSAQTEVQPTQIDGPFAGGKDTVVNINRINTNADGT 410
Db 215 NGGTIKLITSTQNNI-----VDFDLA-----IATDQTV 243
Qy 411 IRVGGFKASITTTNAHLHI-GK-GGVNLSNOAGRSILVENLTGNTITVDGPLRVNNQVGG 468
Db 244 V-----DASSLTNAQTLTNGKIGTICANNKTLQGFNIGSSKT--VLSNGVNAINELVIG 296
Qy 469 YALAGSSANFEKAGTDTKNGTA-----TFNNDISLGRFVNLKYDAHTANFKGIDTG 520
Db 297 NDGAQVFAHDYLI-TRTTNAAOGKIIIFNPVYVNGTTLAAGTNL--GSATNPLAEINFG 353
Qy 521 NGGFNTLDFSGVTDKVNINKLITASTNVAVKFNINELIVKTNGISVGEVTHFSEDIGSQ 580
Db 354 SKGVNVDTVLVNVEGVNL-----YATNITTTDANGVGFENAGGTNI-----VSGTVGGQ 403
Qy 581 --SRINTVRLETGRSLFSGGVKFKGKELVIDEYFYSPWNY-----FDARNIKNVEI 631
Db 404 QGNKFTVALENGTTVKFLNGATFNNGTITAAANSTLQIGNYTADCVASADGTGI--VEF 461
Qy 632 TN-----KLAFGPOGSP-----WGTSKLMENNTLQGN--AVMDYQSFNLTIGD 675
Db 462 VNTGPIVTTLKQAPVNAALKQTVSGPGNVVINEIGNAGNHGAVTDTTAFENSSL--- 518
Qy 676 FINNOGTINYLVRGKVATLSVGNAAAMFNNDIDSAT--GFYKPLIKINSADOLIKNTE 733
Db 519 -----GAVVELPRG--IPENDAGNTMPLTIKSTVGNKTAKGFDVPSVVLGVDSVIADGQ 571
Qy 734 HVLKAKIIGY-----NVSTGTNGISNVNLEP-----TNTTNSARS 858
Db 572 VIGDQNNIVGLGSDNGIIVNATTLVAGISTLNNNGTTLTSGGVPTNPTGTVGLGTGI 631
Qy 762 ---QFKERLAL--YNN-----NNRMDTCVVRVTDIKACG---MAIGDQSVNPNPDN 805
Db 632 GASKFQVTTTDDYNNLGNIIATNATINDGVTVTTGGIAGIGFDGKITLGS---VNGNGN 688
Qy 806 KYK---LIGKAWKNIGISKANGSKISVYVLGNSTPTTENGNTNLP-----TNTTNSARS 858
Db 689 VRADGILNSTSMIGTKANNG--TVTVLGNFAFVGNIGSDTTPVASVFTGSDSAGL 745
Qy 859 ANNALQAAPFAPSATPNLVAINHDF-----GT----- 888
Db 746 QGNIYSQVIDF---GTYNLGIVNSIILGGTTAINGKIDLVNTLTFTFASGTSTWGNNT 801
Qy 889 -IESVFELANRS-----KIDTLYAN-SGAQGRDLLO----- 918
Db 802 SIETTLTLANGNTHGHIIVLBGAQVNTTTTGTIKVODNANANFSGTQVTVTLQGGARFN 861
Qy 919 -TLIDSHDAGYARKMIDAT---SANE---ITKQNTATTTLNNAISLEH----- 961
Db 862 GTLGSPPFAVTSNRFVNSLIRAAQNDYVITRTNNAENVVNTDIANSFPFGAGPVQDNV 921
Qy 962 -----KTSGLQTL-----SLNAMILNSRLVNLRRHTNHIDSFARKLQALKDQKFA 1009
Db 922 TTFVNATNTAAYNLLAKNSANSANFVGAIVTDTSAITNVQLDLAKDQIAQGLNRLGA 981
Qy 1010 LESAAEVLQYFAPKYEKPTVWNAI-----GGTSLNN----- 1042
Db 982 LR-----YLGTPETAEMAGPEAGISAAVAAGDEAIDNVAYGIWAKPFTTDAHQSKKG 1035

Qy 1043 --GSNASLYGTSAGDAYLNGQVEAIVGGFGSYGYSSFNRRANSLSNG-----ANNTFGV 1096
Db 1036 LAGVKATTTGVVIGLDTLANDNL--MIG--AAIGITKTDIKHODYKKGDKTDVNGFSFSL 1091
Qy 1097 YSRIFANQHEFDEFAQG-ALGSDOSSNFKSALLQDLNQSY-----HYLAISAAATRAS 1148
Db 1092 YG---AQOLVKNFFAQGSAIFSLNQVKNKSORYEPFDANGNMSKQIAAGHYDNMTFEGNLT 1148
Qy 1149 YGYDFAFRNALVLKPSGVSYNHLGSTNFKSNSTNOVALKNGSS 1193
Db 1149 VGYDYNAMQGVLV-TPMAGLSYLKSSDENYKETGTT-VANKQVNS 1191
RESULT 13
ID OMPA_RICCN STANDARD; PRT: 2021 AA.
AC Q52657; P95591; P95592; P95593; P95594; Q52667; Q52668; Q52669;
AC Q52670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
antigen) (rOmpA) (rOmp A).
GN OMPA OR RC1273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Croquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.-E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
analysis of restriction fragment length polymorphism of PCR-amplified
DNA of the gene encoding the protein rOmpA";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.-E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
of the outer surface protein rOmpA";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- PTM: GLYCOSYLATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	DR	EMBL; U01028; AAL17405.1; -	
CC	DR	EMBL; AE008674; AAL03811.1; -	
DR	EMBL; U43794; AAB49549.1; -		
DR	EMBL; U43798; AAB49550.1; -		
DR	EMBL; U43806; AAB49551.1; -		
DR	EMBL; U45244; AAB49566.1; -		
DR	EMBL; U46918; AAB86663.1; -		
DR	EMBL; U83440; AAC35176.1; -		
DR	EMBL; U83443; AAC35179.1; -		
DR	EMBL; U83448; AAC35184.1; -		
DR	EMBL; U83453; AAC35189.1; -		
DR	InterPro; IPR003858; rOmpA_rOmpB.		
DR	Pfam; PF02708; rOmpA_rOmpB.1.		
KW	Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;		
KW	Complete proteome.		
FT	SIGNAL	1	POTENTIAL.
FT	CHAIN	39 2021	OUTER MEMBRANE PROTEIN A.
FT	DOMAIN	238 946	THR-RICH.
FT	DOMAIN	1424 1528	THR-RICH.
FT	VARIANT	60 60	N -> NN (IN STRAIN INDIAN TICK TYPHUS).
FT	VARIANT	76 76	R -> H (IN STRAIN INDIAN TICK TYPHUS).
FT	VARIANT	86 137	MISSING (IN STRAIN MI).
FT	VARIANT	126 133	MISSING (IN STRAIN MOROCCAN).
FT	VARIANT	953 954	VT -> II (IN STRAIN INDIAN TICK TYPHUS).
FT	VARIANT	1245 1245	D -> A (IN STRAINS INDIAN TICK TYPHUS, M1 AND MOROCCAN).
FT	VARIANT	1308 1308	N -> H (IN STRAIN MOROCCAN).
FT	VARIANT	1877 1877	M -> I (IN STRAIN INDIAN TICK TYPHUS).
FT	CONFLICT	10 10	Q -> K (IN REF. 1).
FT	CONFLICT	92 92	I -> V (IN REF. 1).
FT	CONFLICT	126 126	V -> I (IN REF. 1).
FT	CONFLICT	137 137	T -> N (IN REF. 1).
FT	CONFLICT	157 157	G -> D (IN REF. 1).
FT	CONFLICT	368 369	IS -> VN (IN REF. 1).
FT	CONFLICT	374 388	KATLGGAIIKATTTK -> LLQVGVVVKANTIN (IN REF. 1).
FT	CONFLICT	640 640	N -> D (IN REF. 1).
FT	CONFLICT	669 669	V -> I (IN REF. 1).
FT	CONFLICT	793 793	N -> D (IN REF. 1).
FT	CONFLICT	803 804	VN -> IS (IN REF. 1).
FT	CONFLICT	809 823	LLRVQGVVVKSNNTIN -> KATLGGAIIKATTTK (IN REF. 1).
FT	CONFLICT	898 898	D -> Y (IN REF. 1).
FT	CONFLICT	908 908	P -> N (IN REF. 1).
FT	CONFLICT	985 985	N -> K (IN REF. 1).
FT	CONFLICT	1009 1009	L -> S (IN REF. 1).
FT	CONFLICT	1013 1013	Y -> S (IN REF. 1).
FT	CONFLICT	1182 1182	K -> Q (IN REF. 1).
FT	CONFLICT	1314 1314	N -> Y (IN REF. 1).
FT	CONFLICT	1451 1451	H -> N (IN REF. 1).
FT	CONFLICT	1624 1624	G -> D (IN REF. 1).
FT	CONFLICT	1628 1628	E -> G (IN REF. 1).
FT	CONFLICT	1872 1872	A -> V (IN REF. 1).
FT	CONFLICT	1875 1875	T -> P (IN REF. 1).
FT	CONFLICT	1878 1879	MS -> LP (IN REF. 1).
FT	CONFLICT	1936 1936	E -> A (IN REF. 1).
FT	CONFLICT	1965 1970	MTAPLP -> ITPPLS (IN REF. 1).
FT	CONFLICT	1997 1997	G -> R (IN REF. 1).
SQL	SEQUENCE	2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;	

Query Match

Best Local Similarity 4.7%; Score 310.5; DB 1; Length 2021;

Matches 302; Conservative 207; Mismatches 533; Indels 463; Gaps 80;

QY	19	LVGALV-----SITPQSHAAFFTTVIIPAIV-----GGIATGTVAGTVSGLLSWGLK 66
DB	740	LGAVIKATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIATFTGDSVTGNTG---- 795
QY	67	QAEANKTPDKPKVVRIOAQ-----KGFNEFPNKEYDLYRS--LLSSKIDGWDGNAAR 120

DB	796	NTNALATVNVGAGLLRVGGVVKSNNTINLTNDASAVTFTNPVVVTGAID---NTGNANN 851
QY	121	HYWVKGGQONKLEVDKMDAVGTYTTLGSLRNFTG-----GDLVNMOKATLRLQGFNG 172
DB	852	-----GIVFTGDSVTGNTGNNTALATISVGAGKATLG-GAIK 890
QY	173	NSFTSYKDSADTRTDFNAKNISIDNFEINNRVSGAGRKASSTVLTLQASEGTTSDK 232
DB	891	ATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIVTGT---DSTV---TGNTG 941
QY	233	NAEISLYDGLATNLASSSVKLMGNVMMGRLOV-----GAYLAPSY----- 273
DB	942	NALATVNVGAGVTLQAGGSLDANNIDFGARSTLEFNGPLDGGGAIPYFYKGAIANGNA 1001
QY	274	STINTSKVTGEVNFENHLLTVGDK-----NAAQAGI 302
DB	1002	ILNVTKLLTA-----YHITGTGVAEINTAGNLFDAIDASAGDVTILNQADIHFRAIDLSAL 1057
QY	303	TANKTNTIGTLDLQMSAGLIITAP--PEGYKDKPNPTPSQSGAKNDKN-----ESAKND 355
DB	1058	VLSNLTGCVGNILLAA--DLVAPGVDEG-----TVVFDGGVNLGNSNVAGAARN 1107
QY	356	KOESSQNSNTQVIPPNSAQKTEVQPTQVIDGPPAGGKDTVVNNINRINTNAD----- 408
DB	1108	IGDVGKNTLLI-----YNAVTTIDDVNLEGIQNVL-----INNADFTSSTAF 1153
QY	409	GTTRVGGFKASLTNAHLHIGKGVNLSNQASGRSLIVENLTGN---ITVDGPLRVN 463
DB	1154	NAGTIQINDATYTTDANNGNLNPAGNIKFAH--ADAQLIOLSSGNDRTITLGANIDPD 1211
QY	464	NO-----VGYALAGSS--ANFEFKAGTDTKNGTATFN-NDISLGRF 502
DB	1212	NDEGIVILNVTACKKLTAGGKTFGGAHKLDIVFKEGEGDTAGTFTNTNIVLDT 1271
QY	503	VNLKVDATAN---FKG-----IDTGN-GGFNTLDFSGVTDKVNINKLITASTNVA--VKN 552
DB	1272	GQLELGATTANVLEFKDAVQLTQTGNIGGF--LDFNAKNGVTYTLNN---NYNVAGTVKN 1325
QY	553	---FNINELIV-----KNGISV-----GRYTFSEDIGSOSRINTVRLGTGRL-- 595
DB	1326	TGCTNGTLLVLGASNLNRVNGIAMLKVAGNWTAK---GCVKIGEIQ-GTGTNLT 1381
QY	596	-----FSGGVKFKGKGLVIDEYFSPWNYFD-----ARNIKVEITNKLAFQ 639
DB	1382	PAHEKLTGSINKTGOALKL-----NFMNGSVGVVGTAAVSGDITTAGATSPAS 1433
QY	640	QCSPWGTSKL-----MFNN---LTLGQNAVMDYSQFSNLTIOGDFINNOGTINYLVR 688
DB	1434	SVNAKGTATLGTTSTFAHTFTNTGAVTLAKGSITSFAK--NVTAT-SFVANSATINF--- 1487
QY	689	GKGVATLSVGNAAAMFNNDID-----SATGFYKPLIKINSAQD-LIKNT 732
DB	1488	-----GNSLA--FNSNTGSGTTLTGANQVYTGTSFTDTLTNTTFDGAARG 1536
QY	733	EHVLKAKIIGYGNVSTGTNGISLVNLEEQFKERLALYNNNNRMBDTCVVVRNTDDIKACGM 792
DB	1537	GNILKS-----GSTLDLSGVSN-----LALVVTATNFD----- 1565
QY	793	AIGDOSMVNN--PDNYKYLIGAKWNIGISKTAGSKISVYVLGNSTPTENGNTNLTPT 850
DB	1566	-----MNNISPDIT-KYTVISA-----ETAGGLK-----PTPKENYKTIINND 1602
QY	851	NTTSNARSANALAQNAFPAQSPATPNLVAIINHDF---GTIESFELANRDKDITLY- 906
DB	1603	RVVDFTFASTLTL-----FAEDIA-----AGVIDEFAFGGPIANIPNANIKKSLELMD 1654
QY	907	ANSAGQGRDLLOT--LLDSDHDGARYKMKIDATS--ANEITQLATATTTLNANISLEHK 962
DB	1655	APNGSDARQAFNFGMLTLPLOEA-----DATHLMQDVVKPSTIAAVNNQVVA--- 1703
QY	963	TSGLOTLSLSNAMIINSLRLVNSRHTNHIDSFARKLQALKDKQKFALESAAEVLYQFAP 1022

Db 1704 ----SNIS-SNITARNMDKVKQAGKGPVSSGDDEM-----DAKFGA-----WISP 1745
Qy 1023 KYEKPTNWMANAIGGTSNLNGSNASLXGTSAGVDAYLNGQVEAIVGFGSGYGSF----- 1078
Db 1746 FVGNAQKMCNSI-----SGYKSDTTGGTIGDPGFSVD--DLVLG-----LAYTRADTDI 1793
Qy 1079 ---NNRANSLNGANN--TNEGVSYSRIFAN-----OHEPDPFAQAGALGSDQSLSN 1123
Db 1794 KLKNNKTGDKKVESNIYSLGLYSVPYENLFVEAIIASYSNDKIRSKSRRIATTLTETVG 1853
Qy 1124 FKSALLQDLNQSYYHAYSAATRASGYDFAFFRNALVLKPSVCVSYNHLGSTNFKSNST 1183
Db 1854 YQTA-----NGKYKSESYTGOLMAGYTY---MMSENLNPLAGLRVSTIKDKSYKETGT 1905
Qy 1184 NOVAL-----KNGSSSOHLFNA--SANVEAR-----YYYGDTSYFYMNAGVLQEFAGVGS 1231
Db 1906 TYQNLTVKGRNYNFDGLGAKYSSNINVAEIVLTPELYAMVDYAFKN----- 1953
1322 NNAASLNTFKVNAARNPLNTHARVMMGGEELKAKEVF-LNKGVVYLHNLSNIGHFASNL 1290
Db 1954 -----KVSADARLOGMTAPLPTNSFKQSKTSFDVGVGVTAKHKMMEYGINYDTNI 2004
Qy 1291 GMY 1294
Db 2005 GSKY 2008

RESULT 14
AIDA_ECOLI
ID AIDA_ECOLI STANDARD; PRT; 1286 AA.
AC Q3135;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin aida-I precursor.
GN AIDA-I.
OS Escherichia coli.
OG Plasmid pIB6.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (O126:H27);
RX MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
"AIDA-I, the adhesin involved in diffuse adherence of the
diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
synthesized via a precursor molecule."
RL Mol. Microbiol. 6:1539-1546(1992).
CC -!- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X65022; CAA46156.1; .
DR PIR; S28634; S28634.
KW Cell adhesion; Signal; Outer membrane; Plasmid.
FT SIGNAL 1 49
FT CHAIN 50 ? ADHESIN AIDA-I.
FT PROPEP ? 1286
SQ SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

Query Match 4.6%; Score 310; DB 1; Length 1286;

Best Local Similarity 21.1%; Pred. No. 6.5e-09;
Matches 304; Conservative 196; Mismatches 568; Indels 376; Gaps 71;
Qy 38 TVIIPATVGGIATGAV---GTVS--GLLSWGLKQ-----AEEANKTPDKPKVRIQAG 87
Db 34 TLLVLAVVSTIGNAFAVNIISGTVSSGGTVSSGGETQIYVSGRGSNATVNSGGTQIVNNGG 93
Qy 88 K----GFNEFPNKEYLDYRLSLSSKIDGWM---DMGNAARHYVWVGQONKLEVDKDA 139
Db 94 KTTATTVNSSGSQNVGTSFATISTIVNSGGIQRVSSGGVASATNLSSGAQN--IYNLGA 151
Qy 140 VGYTTLSSGLRN--PTGGDL--VNMQKATRLQGPNGNSFTYSKDSADRTVRVDENAKNIS 196
Db 152 SNTVIFSGGQTIFSGGTTDSTNISS-----GGQORVSSGGVAGSANTTINSGAQNI- 202
Qy 197 IDNFVEINNRVGSQ-----AGRKASSTVLTLQASEGITSKNAEISLYDGATLINLASSS 250
Db 203 LSEBGLAISTHISISSGQNYIISAGANATETIVNSGGFQRVNSGAVATGTVLSGGTQNVSSGG 262
Qy 251 VKLMGNVMMGRLOVQYAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQAAGIITANKKTI 310
Db 263 SAISTSVYNSGVQTVFAGATVTDITVNSG-----GNQNISSGGIIVSETTVNV 309
Qy 311 -GTLDLWQSAGLNIIPPEGGYKDKPNPTFSQCAKNDKNESAKNDKQESQNNSTQVI 369
Db 310 SGTQNIY-----SGG-----SALSANIKGSOIV 332
Qy 370 NPPNSAQKTEVQP---TQVIDGPFAGGKDTVYN-----INRIN-----TNADGTIRV- 413
Db 333 NSEGTATNLVSDGGYQHIRNGGIASG--TIVNSGVYVNISSGGYAESTINSGGTIRVL 390
Qy 414 --GGFKASLTNAAHLHIGKGVNLSNQASGRSLIVENLTGN--ITVDGRLRVNNOVGGY 469
Db 391 SDGVARGTILNNSGRENVSNGGVSYN-----AMINTGNOVIYSDGEATA----- 435
Qy 470 ALAGSSANFEFKAGTDTKNGTATFNNDISLGRFVNLKVDHAHTANFKGIDTNGCGFNTL-- 527
Db 436 AIVNTSGFORINSQ-----GTAPVQNSVVVTRTVSSAAKPFDA-----EYVSGKQTVYL 485
Qy 528 -----DFSGVTDKVNIN-----KLTASTNVA 549
Db 486 WRGIWYSNFLTAVWMPPTASGANVNLSCRLNFAFAGNVVGTILNQEGROYVYSGATATS 545
Qy 550 VKNFNINELIVKNGISVGEYTHFS--EDIGSQSRIN-TVRLETGTRLSFSG-----GVKF 602
Db 546 TVGNNEGREYVLGGITDGTVLNSGGQLQAVSSGGKASATVINEGGAQFVYDGGQVGTGNI 605
Qy 603 KGEKLVIDE-----FYSPWNYFDARIKKNVEITNKLAFGPGQSPWGTSKLMFNLT 655
Db 606 KNGGTIRVDSGASALNIALSSGGLFTSTGATLPELTMAAL-----SVSQNHASNIV 658
Qy 656 LGQNAVMDYSQFNSLTIOGDFINNOGTVNLVRGGKVATLSVGNAAAMFNNDIDSATGF 715
Db 659 LKNGLLRVF--SGGTATDTVNSAGRLR-IDDGGTNGTNGTTINADGIVAGTINQDGNF 715
Qy 716 YKPLIKINSADLIKNTPE-----HVLK--AKIIGYVNSTGTNGISNVNLEOPKRLAL 769
Db 716 I-----LNLAEYDFETELSGSGVLVKDNTGIMTYAGTLTQAQGVN-----V 757
Qy 770 YNNNRMDTCVVRNTDDIKACGMAIGDSVWNPNDNYKYLIGKAKWNLIGISKTAGSKIS 829
Db 758 KNGGIIFDSAVV-NAD-----MAVQNQAYINTSD--QATINGSVNNGN-SIVINNSIIN 807
Qy 830 VYVLGNST-----PTENG--NTTNLPTNTTSNARSANALQAQAPPAQS 873
Db 808 ----GNTNDADLSFGTAKLLSATVNSGLVNNKNIILNPTK--ESAGNTITVSNYTTGPG 861
Qy 874 ATPNLVAINHDFGTIESFELANRSKIDITLYAN-----SGAQRDLLOTLLIDSH-DAGY 929
Db 862 SVISLGGVLEGDNLSLTDRLVVKNGTSGSDIVVWNEGSGGQTRDGINIISVEGNSDAEF 921
Qy 930 ARKMIDATSAEITKQNLNTATTATTLNNTASLEHKTSGLOT-----LSLSN 973

Db 922 SLKNRVAGAYDTLQKNGESGTDKNGWYL---TSHLPTSDTRQYRPENGSVATNMALAN 978
QY 974 AMILNSRLVNLSSRRHTNHIDSPAKRLQALKDQKFALESAAEVLVQFAPKYEKPTNVWAN 1033
Db 979 SLF-----LMDLNR-----KQFRASDN--TQPESA-----SVWMK 1008
QY 1034 AIGGTS---LNGSN---ASLYGTSGVDAY-----LNGQVEAIVGGFSGYGSFNNR 1081
Db 1009 ITGGISSGKLDQNKTTNQFINQLGGDIYKFAEQLGDFTLGIMGGVANAAGKXTINTY 1068
QY 1082 ANSLNGANNNT---NFGVYSRIFANQHEFDEAGALGSDOSSLN-PKSALLQDLNOSY 1136
Db 1069 S---NKAARNTLDGYSVGIPTWYNGE---NATGLFAETWMOYNWFNASVKGDLGEE 1121
QY 1137 HYLAYSAASTRASGYDEFAFRNALVKPSVGS-----YNHLGS-----TNFKSNST 1183
Db 1122 KYNLGLFASAGGGYNL-----NVHTWTSPEGITGEFWLQPHQAVVMGVTPDTHQEDNGT 1177
QY 1184 -NOVALKNGSSSOHLFNASANVEARYYYGDTSYF--YMNAGVL---QEFARHVGSSNAASL 1237
Db 1178 VVOGAGKNNIQTKAGRASWKVKSLDKDTGRRFRPYIEANWIHNTHREGVKMSDDSQLL 1237
1238 NTFKVNARNPLNTHARVMMGELKAEVFLNGLVYVYLNHLNISIGHFASN-----LGM 1292
Db 1238 -----SGSRN-----QGEIKTGIEGVITQNLVSVNGGVAY-----QAGGHGSNAISGALGI 1282
QY 1293 RYSF 1296
Db 1283 KYSF 1286

RESULT 15
OMP_B_RICPR
ID OMP_B_RICPR STANDARD: PRT: 1643 AA.
AC Q53020: Q53020;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (ScA5) (rOmpB)
DE antigen B [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
OS OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
SC MEDLINE=91045972; PubMed=2122457;
RT Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BREINL;
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompB of Rickettsia prowazekii.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN [4]

PARTIAL SEQUENCE.
RC STRAIN=BREINL;
RX MEDLINE=921114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN [5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hacksadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M37647; AAA26390.1; ALT_INIT.
DR EMBL; AF161079; AAD42234.1; -;
DR EMBL; AJ235273; CAAL5140.1; -;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 Y -> A (IN STRAIN BREINL).
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
FT CONFLICT 178 179 AA -> VC (IN REF. 1).
FT CONFLICT 191 201 T TQAPLPLTGA -> INSRSSSYHLVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735PDF392B6346CC CRC64;
Query Match 4.6%; Score 306; DB 1; Length 1643;
Best Local Similarity 18.6%; Pred. No. 1.5e-08;
Matches 293; Conservative 179; Mismatches 556; Indels 544; Gaps 67;
QY 20 VCALVSIPTQOSHAAFFTTVIIPAIVGGIATGTAAGTAVTSGLLSWGLKQAEAKNTPDKPD 79
Db 109 LGFITNIAQQ----AKFFNFTV-----AAGKILNITGQGI-TVOEASNTINAO 152
QY 80 KVMRIQAGKGFN-----EF-----PNKEYDLYRSLSSKIDGGWDM 115
Db 153 ALTKVHGGAANANDLSGLGSITFAAAPSVLEFNINPTQEAPLTLGANSKIVNGG--- 209
QY 116 GNAARHYWVKGGQONKLEVDMDKDAVGTVTLSGLRNFPTGGDLVDNMOK-----ATLRLGQFN 171
Db 210 -----NGTLNITNGFIQVSDNTFAGIKTINIDCQ 239
QY 172 GNSFTSYKSDADTRTRVDENAKNISIDNFVEINNRVSGAGRKASSTVTLTQASEGITS 230
Db 240 GLMFNSTPDAA-----NTLNQVGGTINFNIGIDGTGK-----LVLVSKNGAAT 283
QY 231 DKNABISLYDG----ATLNASSSVKLMG-----NVWMGRILQ 263

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2002, 11:28:34 ; Search time 46.36 Seconds
(without alignments)
4836.096 Million cell updates/sec

Title: US-09-921-157-3

Perfect score: 6671

Sequence: 1 MEIQQTHRKINRPLVSLV.....HNLISNIGHFASNLGMYSF 1296

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_19.*
- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_page.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6353.5	95.2	1295	2	034110 helicobacte
2	6273	94.0	1290	2	034111 helicobacte
3	6246	93.6	1290	2	09ZHT1 helicobacte
4	6180	92.6	1296	2	09ZHV2 helicobacte
5	6165	92.4	1296	2	09ZHV1 helicobacte
6	6163	92.4	1296	2	09ZHU3 helicobacte
7	6162	92.4	1288	2	09ZUM8 helicobacte
8	6154	92.3	1296	2	09R958 helicobacte
9	6135.5	92.0	1291	2	09ZHU8 helicobacte
10	6128.5	91.9	1291	2	09ZHT8 helicobacte
11	6122	91.8	1296	2	09ZHT7 helicobacte
12	6122	91.8	1296	2	034108 helicobacte
13	6120.5	91.7	1291	2	09LBC1 helicobacte
14	6117.5	91.7	1291	2	09ZHT5 helicobacte
15	6113	91.6	1290	2	096021 helicobacte
16	6112.5	91.6	1291	2	09ZHU1 helicobacte

17	6111.5	91.6	1289	2	093UM6 helicobacte
18	6105.5	91.5	1291	2	09ZHU9 helicobacte
19	6105.5	91.5	1291	2	09ZHT2 helicobacte
20	6101.5	91.5	1291	2	09ZHV3 helicobacte
21	6100.5	91.4	1287	2	093UM9 helicobacte
22	6089.5	91.3	1291	2	09ZHU7 helicobacte
23	6081.5	91.2	1291	2	09ZHU4 helicobacte
24	6080.5	91.1	1291	2	09ZHU6 helicobacte
25	6075.5	91.1	1291	2	09R961 helicobacte
26	6069.5	91.0	1300	2	09ZHU0 helicobacte
27	6064.5	90.9	1291	2	09ZHT9 helicobacte
28	6064.5	90.9	1293	2	09ZHV0 helicobacte
29	6061	90.9	1298	2	09ZHT4 helicobacte
30	6060.5	90.8	1291	2	09R962 helicobacte
31	6033.5	90.4	1287	2	093UM7 helicobacte
32	6012	90.1	1294	2	09R959 helicobacte
33	5862.5	87.9	1303	2	09KJA6 helicobacte
34	5574.5	83.6	1323	2	087018 helicobacte
35	5560	83.3	1324	2	09LBC3 helicobacte
36	5552	83.2	1324	2	09LBC2 helicobacte
37	5539	83.0	1324	2	09LBC7 helicobacte
38	5520	82.7	1328	2	09LBC8 helicobacte
39	5181.5	77.7	1313	2	093UN0 helicobacte
40	5114.5	76.7	1082	2	09R964 helicobacte
41	3929.5	58.9	829	2	09R9F8 helicobacte
42	3926.5	58.9	829	2	09R9G0 helicobacte
43	3918	58.7	826	2	09R9G4 helicobacte
44	3683.5	55.2	839	2	093Q01 helicobacte
45	3434	51.5	866	2	09R9G6 helicobacte

ALIGNMENTS

RESULT 1

ID 034110 PRELIMINARY; PRT; 1295 AA.
AC 034110;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F37;
RX MEDLINE=97339580; PubMed=9196179;
RA Ito Y., Azuma T., Ito S., Miyaji H., Hirai M., Yamazaki Y., Sato F.,
RA Kato T., Kohli Y., Kuriyama M.;
RT "Analysis and typing of the vacA gene from cagA-positive strains of
RT Helicobacter pylori isolated in Japan."
RL J. Clin. Microbiol. 35:1710-1714(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F37;
RX MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori";
RL J. Infect. Dis. 178:1391-1398(1998).
RN [3]
RP SEQUENCE OF 638-752 FROM N.A.
RC STRAIN=JK2-55;
RA Yamaoka Y., Kodama T., Graham D.Y., Kita M., Imanishi J., Kashima K.;
RT "vacA genotypes of Helicobacter pylori in relation to cagA status,
RT cytotoxin production, or clinical outcome."
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF071095; AAC77450.1; -.

DR EMBL: AF042736; AAB99828.1; -.
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca; 1.
SQ SEQUENCE 1295 AA; 139808 MW; B4B7F7AEF7901CB8 CRC64;

Query Match 95.2%; Score 6353.5; DB 2; Length 1295;
Best Local Similarity 95.2%; Pred. No. 1.7e-235;
Matches 1234; Conservative 28; Mismatches 33; Indels 1; Gaps 1;

QY 1 MEIQOHRKINRPLVSLALVGLVSIIPQOSHAAFFTTVLIIPALVGGIATGTVGSGL 60
DB 1 MEIQOHRKINRPLVSLALVGLVSIIPQOSHAAFFTTVLIIPALVGGIATGTVGSGL 60

QY 61 LSWGLKQAEANKTPDKPDKVRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGNAAR 120
DB 61 LSWGLKQAEANKTPDKPDKVRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGNAAR 120

121 HYWVGQQNKLEVDMDKDAVGTYTLSGLRNPFTGGDLVDNNMOKATLRIGQNGNSFTSYKD 180
DB 121 HYWVGQQNKLEVDMDKDAVGTYTLSGLRNPFTGGDLVDNNMOKATLRIGQNGNSFTSYKD 180

QY 181 SADRTTRVDNNAKISIDNVEINNRVSGAGRKASSTVLTLOASEGITSDKNAEISLYD 240
DB 181 SADRTTRVDNNAKISIDNVEINNRVSGAGRKASSTVLTLOASEGITSDKNAEISLYD 240

QY 241 GATLNLAASSVYKLMGNVWMLQVVGAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
DB 241 GATLNLAASSVYKLMGNVWMLQVVGAYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300

QY 301 GIANKNTNIGTLDWOSAGLNIIPAPGGYKDKPNPTPSOGAKNDKNSAKNDKQESS 360
DB 301 GIANKNTNIGTLDWOSAGLNIIPAPGGYKDKPNPTPSOGAKNDKNSAKNDKQESS 360

QY 361 QNNSNTQVIPPNSAOKTEQPTQVIDGPRAGGKDTVVNINRINTNADGTRVGGFKASL 420
DB 361 QNNSNTQVIPPNSAOKTEQPTQVIDGPRAGGKDTVVNINRINTNADGTRVGGFKASL 420

QY 421 TTNAHLHGKGGVNLNSQASGRSLIVENLTGNTVDGRLRVNNOVGAYLAGSSANFEF 480
DB 421 TTNAHLHGKGGVNLNSQASGRSLIVENLTGNTVDGRLRVNNOVGAYLAGSSANFEF 480

QY 481 KAGTDTKNGTATFNNDISLGRFVNLKVDAHTANPKGIDTNGGNTLDFSGVTDKVNINK 540
DB 481 KAGTDTKNGTATFNNDISLGRFVNLKVDAHTANPKGIDTNGGNTLDFSGVTDKVNINK 540

541 LITASTWVAVKFNINELIVKTNIGISVGEYTHFSEDIGSOSRINTVRLTGTSLFSGGV 600
DB 541 LITASTWVAVKFNINELIVKTNIGISVGEYTHFSEDIGSOSRINTVRLTGTSLFSGGV 600

QY 601 KFKGGERLVIDEFYSPWNTFDARNIKNVEITNKLAFQPGQSPMGTSKLMFNLLTQNA 660
DB 601 KFKGGERLVIDEFYSPWNTFDARNIKNVEITNKLAFQPGQSPMGTSKLMFNLLTQNA 660

QY 661 VMDYSQSNLTIOGDFTNNGTINLVRGKVAATPSVGNAAAMFNNDIDSATGYFKPLI 720
DB 661 VMDYSQSNLTIOGDFTNNGTINLVRGKVAATPSVGNAAAMFNNDIDSATGYFKPLI 720

QY 721 KINSAQDLIKNTEHVLKAKTIGVGNVSTGTNGISNVNLEEOPKRLALYNNNRMDTCV 780
DB 721 KINSAQDLIKNTEHVLKAKTIGVGNVSTGTNGISNVNLEEOPKRLALYNNNRMDTCV 780

QY 781 VRNPTDDIKACGMATIGDSVMNPNPNKYLIKAKWNIGISKANGSKISVYVLGNSTPTE 840
DB 781 VRNPTDDIKACGMATIGDSVMNPNPNKYLIKAKWNIGISKANGSKISVYVLGNSTPTE 840

QY 841 NGGNTNLPNTTNSARSANNAQAQAPPAQPSATPNLVAIQNHDFGTIESVFELANRSK 900
DB 841 NGGNTNLPNTTNSARSANNAQAQAPPAQPSATPNLVAIQNHDFGTIESVFELANRSK 900

QY 901 DIDTLYNSGAQGRDLLQTLIDSHDAGYAKMKIDATSAEITKQATATTTLNNAISLE 960
DB 901 DIDTLYNSGAQGRDLLQTLIDSHDAGYAKMKIDATSAEITKQATATTTLNNAISLE 960

990 DIDTLYTHSQAQGRDLLQTLIDSHDAGYAKQMKIDNTSTGTEITKQALNAATDALNNVASLE 959

QY 961 HKTSGQLTSLSNAMILNSRLVNLRRHTNHIDSFARLRQALDKQKPFASLESAAEVLQF 1020
DB 960 HKTSGQLTSLSNAMILNSRLVNLRRHTNHINSFAQRRLQALKGQRFASLESAAEVLQF 1019

QY 1021 APKYERTNWNANAIGTSLNNGSNASLYGTSAGVDAYLNGOVEAIVGGFGSYGYSFNN 1080
DB 1020 APKYERTNWNANAIGGASLNGSNASLYGTSAGVDAYLNGNVEAIVGGFGSYGYSFSN 1079

QY 1081 RANSLGANNNTFGVYSRIFANQHEFDFAQAGALGSDQSLNFKSALLQDLNQSYHYLA 1140
DB 1080 QANSLGANNNTFGVYSRIFANQHEFDFAQAGALGSDQSLNFKSALLQDLNQSYHYLA 1139

QY 1141 YSAATRASGYDYFAFFRNALVLKPSGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1200
DB 1140 YSAATRASGYDYFAFFRNALVLKPSGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1199

QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFAGHVSNNASLNTFKVNAARNPLNTHARVMGGE 1260
DB 1200 SANVEARYYYGDTSYFYMNAGVLQEFAGHVSNNASLNTFKVNAARNPLNTHARVMGGE 1259

QY 1261 LKLAKEVFLNLGVVYLNHLISNIGHFASNLGMYSF 1296
DB 1260 LQLAKEVFLNLGVVYLNHLISNIGHFASNLGMYSF 1295

RESULT 2

034111 PRELIMINARY; PRT; 1290 AA.
AC 034111;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
NCBI_TaxID=210;
RN [1]
RP SEQUENCE OF 512-755 FROM N.A.
RC STRAIN=F79;
RX MEDLINE=97339580; PubMed=9196179;
RA Ito Y., Azuma T., Ito S., Miyaji H., Hirai M., Yamazaki Y., Sato F.,
RA Kato T., Kohli Y., Kuriyama M.;
RT "Analysis and typing of the vacA gene from cagA-positive strains of
RT Helicobacter pylori isolated in Japan";
RL J. Clin. Microbiol. 35:1710-1714(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F79;
RX MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori";
RL J. Infect. Dis. 178:1391-1398(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=F79;
RX Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF071097; AAC77452.1; -.
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca; 1.
SQ SEQUENCE 1290 AA; 139322 MW; CDA478D88AF30E67 CRC64;

Query Match 94.0%; Score 6273; DB 2; Length 1290;
Best Local Similarity 94.0%; Pred. No. 2e-232;
Matches 1218; Conservative 39; Mismatches 33; Indels 6; Gaps 2;

QY 1 MEIQOQTHRKINRPLVSLALVGALVSITPQOSHAAFTTIIIPAIVGGIATGTAVGTVSGL 60
DB 1 MEIQOQTHRKINRPLVSLALVGALVSITPQOSHAAFTTIIIPAIVGGIATGTAVGTVSGL 60
QY 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
DB 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
QY 121 HYWVGGOQNKLEVDKMDKAVGTYTTLISGLRNFSGDLDVNMQKATLRLGQFNQNSFTSYKD 180
DB 121 HYWVGGOQNKLEVDKMDKAVGTYTTLISGLRNFSGDLDVNMQKATLRLGQFNQNSFTSYKD 180
QY 181 SADRTTRVDFNAKNISIDNPFVEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
DB 181 SADRTTRVDFNAKNISIDNPFVEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
QY 241 GATNLASSSVKLMGNVWGRLOYVGCAYLAPSYSTINTSKVTGEVNFHLLTVGDKNAQA 300
DB 241 GATNLASSSVKLMGNVWGRLOYVGCAYLAPSYSTINTSKVTGEVNFHLLTVGDKNAQA 300
QY 301 GIANKKTNIPTLDLWQAGLNIIPAPEGGYKDKPNNTPSQSGAKNKESAKNDKQESS 360
DB 301 GIANKKTNIPTLDLWQAGLNIIPAPEGGYKDKPNNTPSQSGAKNKESAKNDKQESS 360
QY 361 QNNSNTQVINPPNSAQKTEVQPTQVIDGPFAGKDTVVNINRINTNADGTIRVGGFKASL 420
DB 361 QNNSNTQVINPPNSAQKTEVQPTQVIDGPFAGKDTVVNINRINTNADGTIRVGGFKASL 420
QY 421 TTNAAHLHIGKGVNLSNOAGRSLLVENLTGNTITVDGPLRVNNOVGGYALAGSSANFEF 480
DB 421 TTNAAHLHIGKGVNLSNOAGRSLLVENLTGNTITVDGPLRVNNOVGGYALAGSSANFEF 480
QY 481 KAGTDTKNGTATFNNDISLGRVNLKVDHAHTAFNGIDTNGGFTNLDPSGVTDKVNINK 540
DB 481 KAGTDTKNGTATFNNDISLGRVNLKVDHAHTAFNGIDTNGGFTNLDPSGVTDKVNINK 540
QY 541 LITASTNVAKVFNFNINELIVKTVSGVGYTHFSDIGSQSRINTVRLTGTSLFSGGV 600
DB 541 LITASTNVAKVFNFNINELIVKTVSGVGYTHFSDIGSQSRINTVRLTGTSLFSGGV 600
QY 601 KFKGGEKLVDFEYKSPWNYFARINIKVNEITNKLAFGPGQSPWGTSKLMFNLLTGQNA 660
DB 601 KFKGGEKLVDFEYKSPWNYFARINIKVNEITNKLAFGPGQSPWGTSKLMFNLLTGQNA 660
QY 661 VMDYSQFSLTIQGGFINNOGNTINLVRGKVKATLSVGNAAAMFNNDIDSATGYKPLI 720
DB 661 VMDYSQFSLTIQGGFINNOGNTINLVRGKVKATLSVGNAAAMFNNDIDSATGYKPLI 720
QY 721 KINSADLLIKNTEHVLKAKIIGYGNVSTGTNGISNVNLEEDQFKERLALYNNNRMDTCV 780
DB 721 KINSADLLIKNTEHVLKAKIIGYGNVSTGTNGISNVNLEEDQFKERLALYNNNRMDTCV 780
QY 781 VRNTDDIKACGMAIGDOSMVNPNPNKYKYLIGKAWKNIGISKTANGSKISVYVLGNSTPTE 840
DB 781 VRNTDDIKACGMAIGDOSMVNPNPNKYKYLIGKAWKNIGISKTANGSKISVYVLGNSTPTE 840
QY 841 NGGNTTNLTNTTNSARSAANALQAPPAQPSATPNLVAIINQHDFTGTTIESVFELANRSK 900
DB 841 NGGNTTNLTNTTNSARSAANALQAPPAQPSATPNLVAIINQHDFTGTTIESVFELANRSK 900
QY 901 DIDTLYANSAGCROLLQTLTLLDSHDAGYARMKIDATSANETKQLNTATTLNNTIASLE 960
DB 901 DIDTLYANSAGCROLLQTLTLLDSHDAGYARMKIDATSANETKQLNTATTLNNTIASLE 960
QY 961 HKTSGLOTLSLNSAMTNSLRNHLNRRHTNHIDSFAPRLQALKDQKFALESAAEVLQYF 1020
DB 961 HKTSGLOTLSLNSAMTNSLRNHLNRRHTNHIDSFAPRLQALKDQKFALESAAEVLQYF 1020
QY 1021 APKYEKPTVWANAIGGTSILNKSASLYCTGAGVDAYLNGQVEAIVGGFGSYGSFNN 1080
DB 1021 APKYEKPTVWANAIGGTSILNKSASLYCTGAGVDAYLNGQVEAIVGGFGSYGSFNN 1080
QY 1081 RANSLNSGANNTNFGVYSRIFANQHEFDEFAOAGALGSDQSSLNFKSALLQDLNQSYHYLA 1140

DB 1075 QANSLNSGANNTNFGVYSRIFANQHEFDEFAOAGALGSDQSSLNFKSALLQDLNQSYHYLA 1134
QY 1141 YSAATRASGYDPAFFRNALVLKPSVGVSYNHLGTSNFKSNSTNOVALKNGSSOHLFNA 1200
DB 1135 YSAATRASGYDPAFFRNALVLKPSVGVSYNHLGTSNFKSNSTNOVALKNGSSOHLFNA 1194
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFPAHFVHGVSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
DB 1195 SANVEARYYYGDTSYFYMNAGVLQEFPAHFVHGVSNNAASLNTFKVNAARNPLNTHARVMMGGE 1254
QY 1261 LKLAKEVFLNLGVVYLNHLNISNIGHFASNLGMRYSF 1296
DB 1255 LKLAKEVFLNLGVVYLNHLNISNIGHFASNLGMRYSF 1290
RESULT 3
Q92HT1 PRELIMINARY; PRT; 1290 AA.
AC Q92HT1;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11637;
RX MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y., Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and noncytotoxic Helicobacter pylori";
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL; AF049653; AAD04290.1; -
DR InterPro; IPR003842; VACA.
DR Pfam; PF02691; Vaca; 1.
SQ SEQUENCE 1290 AA; 139551 MW; 3DCE42C519352541 CRC64;
Query Match 93.6%; Score 6246; DB 2; Length 1290;
Best Local Similarity 93.4%; Pred. No. 2.2e-231;
Matches 1211; Conservative 40; Mismatches 39; Indels 6; Gaps 2;
QY 1 MEIQOQTHRKINRPLVSLALVGALVSITPQOSHAAFTTIIIPAIVGGIATGTAVGTVSGL 60
DB 1 MEIQOQTHRKINRPLVSLALVGALVSITPQOSHAAFTTIIIPAIVGGIATGTAVGTVSGL 60
QY 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
DB 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
QY 121 HYWVGGOQNKLEVDKMDKAVGTYTTLISGLRNFSGDLDVNMQKATLRLGQFNQNSFTSYKD 180
DB 121 HYWVGGOQNKLEVDKMDKAVGTYTTLISGLRNFSGDLDVNMQKATLRLGQFNQNSFTSYKD 180
QY 181 SADRTTRVDFNAKNISIDNPFVEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
DB 181 SADRTTRVDFNAKNISIDNPFVEINNRVSGAGRKASSTVLTLOASEGITSKNAEISLYD 240
QY 241 GATNLASSSVKLMGNVWGRLOYVGCAYLAPSYSTINTSKVTGEVNFHLLTVGDKNAQA 300
DB 241 GATNLASSSVKLMGNVWGRLOYVGCAYLAPSYSTINTSKVTGEVNFHLLTVGDKNAQA 300
QY 301 GIANKKTNIPTLDLWQAGLNIIPAPEGGYKDKPNNTPSQSGAKNKESAKNDKQESS 360
DB 301 GIANKKTNIPTLDLWQAGLNIIPAPEGGYKDKPNNTPSQSGAKNKESAKNDKQESS 360
QY 361 QNNSNTQVINPPNSAQKTEVQPTQVIDGPFAGKDTVVNINRINTNADGTIRVGGFKASL 420
DB 361 QNNSNTQVINPPNSAQKTEVQPTQVIDGPFAGKDTVVNINRINTNADGTIRVGGFKASL 420

Db	356	QNNNTQVNPNSAQKTEIQPTQVINGPFAGGKDTVVNINRINTNADGTTTRVGGYKASL	415
Qy	421	TTNAHLHLHGKGGVNLNSQASGRSLIVENLTGNTITVDGPLRVNNOVGGLAGSSANFEF	480
Db	416	TTNAHLHLHGKGGVNLNSQASGRSLIVENLTGNTITVDGPLRVNNOVGGLAGSSANFEF	475
Qy	481	KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTGNNGFNTLDFSGVTDKVNINK	540
Db	476	KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTGNNGFNTLDFSGVTDKVNINK	535
Qy	541	LITASTNVAVKFNFINELIVKTNISVGEYTHFSEIDIGSQSRINTVRLTSTRSLFSGGV	600
Db	536	LITASTNVAVKFNFINELIVKTNISVGEYTHFSEIDIGSQSRINTVRLTSTRSLFSGGV	595
Qy	601	KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPWGTSKLMFNLLTGONA	660
Db	596	KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPWGTSKLMFNLLTGONA	655
Qy	661	VMDYSQFSLTIQDGFINNQGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI	720
Db	656	VMDYSQFSLTIQDGFINNQGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI	715
Qy	721	KINGAODLIKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNNRMDTCV	780
Db	716	KINGAODLIKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNNRMDTCV	775
Qy	781	VRNTDDIKACGMAIGDQSMVNPONKYKLGKAMKNTGISKANGSKTSVYLLGNSTPTE	840
Db	776	VRNTDDIKACGMAIGDQSMVNPONKYKLGKAMKNTGISKANGSKTSVYLLGNSTPTE	835
Qy	841	NGGNTNLTPTNTTNSARSANALAQNAFFPAQSPATPNLVAIQNDHFGTIESVFELANRSK	900
Db	836	NGGNTNLTPTNTTNSARSANALYKNAPFAH-SATPNLVAIQNDHFGTIESVFELANRSK	894
Qy	901	DIDTLYNSGAGGRDLQTLTLLIDSHDAGYARKMIDATSNANETKQLTATTTLNNIASLE	960
Db	895	DIDTLYNSGAGGRDLQTLTLLIDSHDAGYARQMDTNTSTGEITKQLNAATTTLNNIASLE	954
Qy	961	HKTSGLQTLTSLNAMLNRLVNLRSRHTNHDISFAKRLQALKDQKPALESAAEVLYQF	1020
Db	955	HKTSGLQTLTSLNAMLNRLVNLRSRHTNHDISFAKRLQALKDQKPALESAAEVLYQF	1014
Qy	1021	APKYEKPTNVWANAIGTSTLANGSNASLYGTSGVDAYLNGQVEAIVGGFGSYGSFNN	1080
Db	1015	APKYEKPTNVWANAIGTSTLANGSNASLYGTSGVDAYLNGQVEAIVGGFGSYGSFNN	1074
Qy	1081	RANSLNSGANTNFGVYSRIPANQHEFDFAQAGALGSDQSSLNFKSALLQDLNOSYHYLA	1140
Db	1075	RANSLNSGANTNFGVYSRIPANQHEFDFAQAGALGSDQSSLNFKSALLQDLNOSYHYLA	1134
Qy	1141	YSAATRASGYDFAFFRNALVKPSGVSYVNLGSTNFKSNSTNOVALKNGSSSOHLFNA	1200
Db	1135	YSAATRASGYDFAFFRNALVKPSGVSYVNLGSTNFKSNSTNOVALKNGSSSOHLFNA	1194
Qy	1201	SANVEARYYYGDTSYFYMNAGVLQEFARHVGSNNAASLNTKPVNAARPLNTHARVMMGGE	1260
Db	1195	SANVEARYYYGDTSYFYMNAGVLQEFARHVGSNNAASLNTKPVNAARPLNTHARVMMGGE	1254
Qy	1261	LKAKEVFLNLGVYLLHNLISNIGHFASNLGMRYSF	1296
Db	1255	LQLAKEVFLNLGVYLLHNLISNIGHFASNLGMRYSF	1290
RESULT	4		
Q92HV2			
ID	Q92HV2	PRELIMINARY;	PRT; 1296 AA.
AC	Q92HV2;		
DT	01-MAY-1999	(Tremblrel. 10, Created)	
DT	01-MAY-1999	(Tremblrel. 10, Last sequence update)	
DE	01-DEC-2001	(Tremblrel. 19, Last annotation update)	
DE	VACUOLATING	CYTOTOXIN.	
GN	VACA.		

OS	Helicobacter pylori (Campylobacter pylori).		
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;		
OC	Helicobacter.		
OX	NCBI_TaxID=210;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=F35;		
RX	MEDLINE=98453456; PubMed=9780260;		
RA	Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,		
RA	Kuriyama M.;		
RT	"Full-length sequence analysis of the vacA gene from cytotoxic and		
RT	noncytotoxic Helicobacter pylori.";		
RT	J. Infect. Dis. 178:1391-1398(1998).		
DR	EMBL: AF049625; AAD04264.1; -;		
DR	InterPro: IPR003842; Vaca.		
DR	Pfam: PF02691; Vaca; 1.		
SQ	SEQUENCE 1296 AA; 139900 MW; A5F86C6BF6C37C9C CRC64;		
Query Match 92.6%; Score 6180; DB 2; Length 1296;			
Best Local Similarity 91.8%; Pred. No. 7.3e-229;			
Matches 1190; Conservative 52; Mismatches 54; Indels 0; Gaps 0;			
Qy	1	MEIOOTHRKINRPLVSLALVGLVSIITPQSHAAFFFTTVIIPAIIVGGIATGTAAGTVSGL	60
Db	1	MELQOOTHKINRPLVSLALVGLVSIITPQSHAAFFFTTVIIPAIIVGGIATGTAAGTVSGL	60
Qy	61	LSWGLKQAEANKTPDKVWRIQAGKGFNFEPNKEVDLYRSLSSKIDGDMGNAAAR	120
Db	61	LGWGLKQAEANKTPDKVWRIQAGKGFNFEPNKEVDLYRSLSSKIDGDMGNAAAR	120
Qy	121	HYWYKGGQNKLEVDMDKDAVGTYTLGSLRNFTGGDLVNMOKATLRLQGFNGNSFTSYKD	180
Db	121	HYWYKGGQNKLEVDMDKDAVGTYKLSGLRNFTGGDLVNMOKATLRLQGFNGNSFTSFKD	180
Qy	181	SADTRTRVDFNAKISIDNFVEINNRVSGAGRKASSTVLTQLASEGITSOKNAEISLYD	240
Db	181	SADTRTRVDFNAKISIDNFLEINNRVSGAGRKASSTVLTQLASEGITSOKNAEISLYD	240
Qy	241	GATNLASSVVKLMGVNMGRLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA	300
Db	241	GATNLASSVVKLMGVNMGRLQVVGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA	300
Qy	301	GIIAKKNTNIGTLWQASAGLNIAPPEGGYKDKPNNTPSQSGAKNDKNEAKNDKQESS	360
Db	301	GIIASKKTYICTLWQASAGLNIAPPEGGYKDKPNNTPSQSGAKNDKNEAKNDKQESS	360
Qy	361	QNNSTQVNPNSAQKTEVQPTQVIDGPPAGGKDTVVNINRINTNADGTTTRVGGYKASL	420
Db	361	QNNSTQVNPNSAQKTEIQPTQVIDGPPAGGKDTVVNINRINTNADGTTTRVGGYKASL	420
Qy	421	TTNAHLHLHGKGGVNLNSQASGRSLIVENLTGNTITVDGPLRVNNOVGGLAGSSANFEF	480
Db	421	TTNAHLHLHGKGGVNLNSQASGRSLIVENLTGNTITVDGPLRVNNOVGGLAGSSANFEF	480
Qy	481	KAGDTKNGTATFNNDISLGRFVNLKVDAHTANFKGIDTGNNGFNTLDFSGVTDKVNINK	540
Db	481	KAGVDTKNGTATFNNDISLGRFVNLKASAHTVNFKIDTGNNGFNTLDFSGVTDKVNINK	540
Qy	541	LITASTNVAVKFNFINELIVKTNISVGEYTHFSEIDIGSQSRINTVRLTSTRSLFSGGV	600
Db	541	LITASTNVAVKFNFINELIVKTNISVGEYTHFSEIDIGSQSRINTVRLTSTRSLFSGGV	600
Qy	601	KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFPGQSPWGTSKLMFNLLTGONA	660
Db	601	KFKGGEKLVINDFYAPWNYFDARNIKNVEITNKLAFPGQSPWGTSKLMFNLLTGONA	660
Qy	661	VMDYSQFSLTIQDGFINNQGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI	720
Db	661	VMDYSQFSLTVIQGFINNQGTINYLVRGGNIETLSVGNAAAMFNNDIDSATGFYKPLI	720
Qy	721	KINGAODLIKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNNRMDTCV	780

Db	1	MELOQTHRKINRPLVLSLALVGLLVSTTPQKSHAAFFTTVITPAIVGGTATGTVGTVSGL	60
Qy	61	LSWGLQAEAEANKTPDKPVWRIQAGKFNFPKNKEDLYBSLLSSKIDGGWDGNAAR	120
Db	61	LGWGLQAEAEANKTPDKPVWRIQAGKFNFPKNKEDLYBSLLSSKIDGGWDGNAAR	120
Qy	121	HYWYGGQONKLEVDKMDAVGTYTTLGSLRNFETGGDLVDNMOKATURLQOFNGSNFTSYKD	180
Db	121	HYWYGGQONKLEVDKMDAVGTYTTLGSLRNFETGGDLVDNMOKATURLQOFNGSNFTSPKD	180
Qy	181	SADRTTRVDFNAKNISIDNFEVNNRNVSGAGRKASSTVLTLOASEGITSDKNAEISLYD	240
Db	181	SADRTTRVDFNAKNILIDNFEVNNRNVSGAGRKASSTVLTLOASEGITSDKNAEISLYD	240
Qy	241	GATLNILASSSVKLMGNVWGRLOQYGAYLAPYSYINTSKYTGVEVNFNHLTVGDKNAQA	300
Db	241	GATLNILASNSVKLMGNVWGRLOQYGAYLAPYSYINTSKVYGEVNFNHLTVGDKNAQA	300
Qy	301	GIANKTKNITIGTLDLWQSLNIIAPPEGGYKDKPNNTPSQGAKNKDNKESAKNDKOESS	360
Db	301	GIASKKTYIGTLDLWQSLNIIAPPEGGYKDKPNNTNSQGAKNKDNKESAKNDKOESS	360
Qy	361	QNSNSTQVINPPNSAOKTEVQPTQVIDGPFAGCKDTVYVNNRINTNADGTTRVGFGFASL	420
Db	361	QNSNSTQVINPPNSGOKTEIQPTQVIDGPFAGCKDTVYVNNRINTNADGTTIKVGYTASL	420
Qy	421	TTNAHLHIGKGVNLSNQAQRSLVENLTGNTIVDGPLRVNNOGVGYALAGSSANPEF	480
Db	421	TTNAANLIGKGVNLSNQAQRSLVENLTGNTIVDGMALVNOGVGYALAGSSANPEF	480
Qy	481	KAGDTKNGTATFNDDISLGRFVNLKVDHAHTANFKGIDTNGGFGNTLDFSGVTOKVINIK	540
Db	481	KAGVDTKNGTIAFNNTISLGRFVNLKASHTVNEKIDTNGGFGNTLDFSGVTNKVNLINK	540
Qy	541	LITASTNVAVKFNINIELIVKNGISVGEYTHFSEDISQSQRINTVRLETGTRSLFSGGV	600
Db	541	LITASTNVAIKFNINIELLVKTNKNGISVGEYTNFSEDISQNRINTVRLETGTRSYSGGV	600
Qy	601	KFKGGEKLVIDEYYSPPWYFQARNIKVNEITNKLAFGQGPSPGCTSKLMFNNTLQONA	660
Db	601	KFKGGEKLVIDEYYPWYFQARNIKVNEITNKLAFGQGPSPGCTSKLMFNNTLQGNA	660
Qy	661	VMDYSQFSLTIQGFINNQGTINYLVRGCKVATILSVGNAAMFNNDIDSATGYFKPLI	720
Db	661	VMDYSQFSNVTIOGFINNQGTINYLVRGNETILSVGNAAMVSPNDIDSATGYFKPLI	720
Qy	721	KINSADOLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEBOFKERLALYNNNRMDTCV	780
Db	721	KINSADOLIKNKEHVLLKAKIIGYENASLGTNSISNANLIEQFNERLALYNNNRMDTCV	780
Qy	781	VRNTDDIKACGMAIGDQSMVNPNDNYKYLIGKAWKNIGISTANGSKTSVYVLGNSPTE	840
Db	781	VRNTDDIKACGMAIGDQAMVNPNDNYKYLIGKAWKNIGISTANGSKTSVRYLGNATPAE	840
Qy	841	NGGNTNLPNTTNSNARSANNAQAAPQSPATPNLVAINQHDFGTIESVFELANRSK	900
Db	841	NGGNTNLPNTNARFASALIKNAFQATNPATPNLVAINKHDFGTIESVFELANRSK	900
Qy	901	DIDTLYANSQAQRDLQTLTLLIDSHDAGYARKMIDATSANETIKOLNTATTTLANNIASLE	960
Db	901	DIDTLYAHSQAQRDLQTLTLLIDSHDAGYARTMIDATSANETIKOLNTATDALNNIASLE	960
Qy	961	HKTSGLQTLSLSNAMTILNSRLYNLSRRHTNHIDSEAKRLQALKDQKFALESAAEVLQOF	1020
Db	961	HKTSGLQTLSLSNAMTILNSRLYNLSRRHTNHIDSEAFRLQALKDQKFALESAAEVLQOF	1020
Qy	1021	APKYEXPTNVWANAIGGTSLLNNGSNASLYGTGSAGVDAYLNGOVEAIVGGFGSGYSGFNN	1080
Db	1021	APKYEXPTNVWANAIGGALSNGGNTSLTSGTGSAGVDAYLNEKVEAIVGGFGSGYSGFNN	1080
Qy	1081	RANSLNSGANTNFGVYSRIFANQHEFDEFAOGALGSDQSSLSNFKSALLQDLNOSHYLA	1140
Db	1081	QANSLNSGANTNFGVYSRIFANRHEFDEFAOGVSGDQSSLSNFKSALLRDLNOSYNLA	1140

QY 1 MEIQOTHRKINRPLVSLALVCALVSITPOOSHAAFFTTVIIPAIVGGIATGTAVGTVSGL 60
|||::|||||:|||||:


```
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AFN4769;
RA Ji X.H., Rappuoli R., Telford J.L.;
RT "Functional analysis of chimeric mutants of the helicobacter pylori
RL vacA gene.";
DR EMBL: AF191643; AAK56858.1; -;
SQ SEQUENCE 1288 AA; 139109 MW; 524ACC3524F0CA3 CRC64;

Query Match 92.4%; Score 6162; DB 2; Length 1286;
Best Local Similarity 91.7%; Pred. No. 3.6e-228;
Matches 1189; Conservative 52; Mismatches 47; Indels 8; Gaps 2;

QY 1 MEIQOTHRKINRPLVSLALVGVLSITPQOASHAAFTTVIIPAIYVGGIATGAVGTVSG 60
DB 1 MEIQOTHRKMNRPVSLVLAGALISAIPQESHAAFETTVIIPAIYVGGIATGAVGTVSG 60
61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKKEYDLYBSLLSSKIDGGWDMGNAAR 120
61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKKEYDLYBSLLSSKIDGGWDMGNAAR 120
121 HYWKGGOQNKLEVDKMDAVGTYTTLGLRNFTGGDLVNMQKATLRLGQFNGNSFTSYKD 180
121 HYWKGGOQNKLEVDKMDAVGTYTTLGLRNFTGGDLVNMQKATLRLGQFNGNSFTSYKD 180
181 SADRTTRVDFNAKNTSIDNFVNNRVSAGKAKASSVTLTQASEGITSKNAEISLYD 240
181 SADRTTRVDFNAKNTSIDNFVNNRVSAGKAKASSVTLTQASEGITSKNAEISLYD 240
241 GATNLASSVKLMGNVWMLQYVGYLAPSYTINTSKVTGEVNFHNLTVGDKNAQA 300
241 GATNLASSVKLMGNVWMLQYVGYLAPSYTINTSKVTGEVNFHNLTVGDKNAQA 300
301 GIANKTKHIGTLDLQWAGLNIAPPPEGKYDKNPNTFSSQAGKNDKNAESAKNDQESS 360
301 GIANKTKHIGTLDLQWAGLNIAPPPEGKYDKNPNTFSSQAGKNDKNAESAKNDQESS 360
361 QNNSNTQVINPNNSAKQTEVQTVIDGPFAGGKDTVNVNINRINTNADGTVIRVGGFKASL 420
361 QNNSNTQVINPNNSAKQTEVQTVIDGPFAGGKDTVNVNINRINTNADGTVIRVGGFKASL 420
355 -NNSNTEVINPNNTQKTEPETEQVIDGPFAGGKDTVNVNINRINTNADGTVIRVGGFKASL 413
355 -NNSNTEVINPNNTQKTEPETEQVIDGPFAGGKDTVNVNINRINTNADGTVIRVGGFKASL 413
421 TTNAHLHIGKGVNLNQASGRSLIVENLTGNTITVDGPLRVNNOVGGYALAGSSANPEF 480
421 TTNAHLHIGKGVNLNQASGRSLIVENLTGNTITVDGPLRVNNOVGGYALAGSSANPEF 480
414 TTNAHLHIGKGVNLNQASGRSLIVENLTGNTITVDGPLRVNNOVGGYALAGSSANPEF 473
414 TTNAHLHIGKGVNLNQASGRSLIVENLTGNTITVDGPLRVNNOVGGYALAGSSANPEF 473
481 KAGTDTKNGTATFNNDISLGRVNLKVDAAHTANFKGIDTNGGNTLDFSGVTDKVNINK 540
481 KAGTDTKNGTATFNNDISLGRVNLKVDAAHTANFKGIDTNGGNTLDFSGVTDKVNINK 540
474 KAGVDTKNGTATFNNDISLGRVNLKVDAAHTANFKGIDTNGGNTLDFSGVTDKVNINK 533
474 KAGVDTKNGTATFNNDISLGRVNLKVDAAHTANFKGIDTNGGNTLDFSGVTDKVNINK 533
541 LITASTNVAVKFNINELLIVKNGISVGEYTHFSEIDGSQSRINIVRLTGRSRFSGGV 600
541 LITASTNVAVKFNINELLIVKNGISVGEYTHFSEIDGSQSRINIVRLTGRSRFSGGV 600
534 LITASTNVAVKFNINELLIVKNGISVGEYTHFSEIDGSQSRINIVRLTGRSRFSGGV 593
534 LITASTNVAVKFNINELLIVKNGISVGEYTHFSEIDGSQSRINIVRLTGRSRFSGGV 593
601 KFKGGEKLVDFYFSPWNYFDARINKVETINKLAFPGQSPWGTSKLMENNLTLGQNA 660
601 KFKGGEKLVDFYFSPWNYFDARINKVETINKLAFPGQSPWGTSKLMENNLTLGQNA 660
594 KFKGGEKLVDFYFSPWNYFDARINKVETINKLAFPGQSPWGTSKLMENNLTLGQNA 653
594 KFKGGEKLVDFYFSPWNYFDARINKVETINKLAFPGQSPWGTSKLMENNLTLGQNA 653
661 VMDYSQFNLTIQGDFFINNOGTYNLYLRGGKVVATLSVGNAAAMFNNDIDSATGYFKPLI 720
661 VMDYSQFNLTIQGDFFINNOGTYNLYLRGGKVVATLSVGNAAAMFNNDIDSATGYFKPLI 720
654 VMDYSQFNLTIQGDFFINNOGTYNLYLRGGKVVATLSVGNAAAMFNNDIDSATGYFKPLI 713
654 VMDYSQFNLTIQGDFFINNOGTYNLYLRGGKVVATLSVGNAAAMFNNDIDSATGYFKPLI 713
721 KINSADLIKNTHEVLLAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDTCV 780
721 KINSADLIKNTHEVLLAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDTCV 780
714 KINSADLIKNTHEVLLAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDTCV 773
714 KINSADLIKNTHEVLLAKIIGYGNVSTGTNGISVNLLEQFKERLALYNNNRMDTCV 773
781 VNTDODIKACGMAIGDQSMVNNPDNYKYILIGKAWNIGISKTAGSKISVYVLGNSTPTE 840
781 VNTDODIKACGMAIGDQSMVNNPDNYKYILIGKAWNIGISKTAGSKISVYVLGNSTPTE 840
774 VNTDODIKACGMAIGDQSMVNNPDNYKYILIGKAWNIGISKTAGSKISVYVLGNSTPTE 833
774 VNTDODIKACGMAIGDQSMVNNPDNYKYILIGKAWNIGISKTAGSKISVYVLGNSTPTE 833
```

```
QY 841 NGGNTTNLPTNTTSNARSANALQAQNPAPQSPATPNLVAINQHDFGTIESVFELANRSK 900
DB 834 NGGNTTNLPTNTTSNARSANALIKNPAFAH-SATPNLVAINQHDFGTIESVFELANRSK 892
QY 901 DDTLYANSAGAGRDLLQTLTLLSDHDAGYARKMIDATSNETTKQNTATTTLNNTASLE 960
DB 893 DDTLYTHSGAGRDLLQTLTLLSDHDAGYARQMDINTSTGEITKQNAATADLNNTASLE 952
QY 961 HKTSGLOTLSLNAMILNSRLNLSRRHNSIDSFARLQALDKDQKFALESAAEVLQYOF 1020
DB 953 HQSGLOTLSLNAMILNSRLNLSRRHNSIDSFARLQALDKDQKFALESAAEVLQYOF 1012
QY 1021 APKYEKPTNVWNAITGGTSLNNGSNASLYGTSGAGVDAYLNGQVEAIVGGFGSYGSFNN 1080
DB 1013 APKYEKPTNVWNAITGGTSLNNGSNASLYGTSGAGVDAYLNGQVEAIVGGFGSYGSFNN 1072
QY 1081 RANSLNNGANNNTNFGYSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNQSXYHLA 1140
DB 1073 QANSLNNGANNNTNFGYSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNQSXYHLA 1132
QY 1141 YSAATRASGYDFAFPRNALVLKPSVGVNHLGSTNFKSNSTNOVALKNGSSSHLFNA 1200
DB 1133 YSATTRASGYDFAFPRNALVLKPSVGVNHLGSTNFKSNSTNOVALKNGSSSHLFNA 1192
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFYAHVGSNNAASLNTFKVNAARNPLNTHARVMMGE 1260
DB 1193 NANVEARYYYGDTSYFYMNAGVLQEFYAHVGSNNDVSLNTFKVNAARNPLNTHARVMMGE 1252
QY 1261 LKLAKEVFLNLGVVYVHLNLSNIGHFASNLGMRYSF 1296
DB 1253 LKLAKEVFLNLGVVYVHLNLSNIGHFASNLGMRYSF 1288

RESULT 8
QY 841 NGGNTTNLPTNTTSNARSANALQAQNPAPQSPATPNLVAINQHDFGTIESVFELANRSK 900
DB 834 NGGNTTNLPTNTTSNARSANALIKNPAFAH-SATPNLVAINQHDFGTIESVFELANRSK 892
QY 901 DDTLYANSAGAGRDLLQTLTLLSDHDAGYARKMIDATSNETTKQNTATTTLNNTASLE 960
DB 893 DDTLYTHSGAGRDLLQTLTLLSDHDAGYARQMDINTSTGEITKQNAATADLNNTASLE 952
QY 961 HKTSGLOTLSLNAMILNSRLNLSRRHNSIDSFARLQALDKDQKFALESAAEVLQYOF 1020
DB 953 HQSGLOTLSLNAMILNSRLNLSRRHNSIDSFARLQALDKDQKFALESAAEVLQYOF 1012
QY 1021 APKYEKPTNVWNAITGGTSLNNGSNASLYGTSGAGVDAYLNGQVEAIVGGFGSYGSFNN 1080
DB 1013 APKYEKPTNVWNAITGGTSLNNGSNASLYGTSGAGVDAYLNGQVEAIVGGFGSYGSFNN 1072
QY 1081 RANSLNNGANNNTNFGYSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNQSXYHLA 1140
DB 1073 QANSLNNGANNNTNFGYSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNQSXYHLA 1132
QY 1141 YSAATRASGYDFAFPRNALVLKPSVGVNHLGSTNFKSNSTNOVALKNGSSSHLFNA 1200
DB 1133 YSATTRASGYDFAFPRNALVLKPSVGVNHLGSTNFKSNSTNOVALKNGSSSHLFNA 1192
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFYAHVGSNNAASLNTFKVNAARNPLNTHARVMMGE 1260
DB 1193 NANVEARYYYGDTSYFYMNAGVLQEFYAHVGSNNDVSLNTFKVNAARNPLNTHARVMMGE 1252
QY 1261 LKLAKEVFLNLGVVYVHLNLSNIGHFASNLGMRYSF 1296
DB 1253 LKLAKEVFLNLGVVYVHLNLSNIGHFASNLGMRYSF 1288

PRELIMINARY: PRT; 1296 AA.
AC QY 841 NGGNTTNLPTNTTSNARSANALQAQNPAPQSPATPNLVAINQHDFGTIESVFELANRSK 900
DB 834 NGGNTTNLPTNTTSNARSANALIKNPAFAH-SATPNLVAINQHDFGTIESVFELANRSK 892
QY 901 DDTLYANSAGAGRDLLQTLTLLSDHDAGYARKMIDATSNETTKQNTATTTLNNTASLE 960
DB 893 DDTLYTHSGAGRDLLQTLTLLSDHDAGYARQMDINTSTGEITKQNAATADLNNTASLE 952
QY 961 HKTSGLOTLSLNAMILNSRLNLSRRHNSIDSFARLQALDKDQKFALESAAEVLQYOF 1020
DB 953 HQSGLOTLSLNAMILNSRLNLSRRHNSIDSFARLQALDKDQKFALESAAEVLQYOF 1012
QY 1021 APKYEKPTNVWNAITGGTSLNNGSNASLYGTSGAGVDAYLNGQVEAIVGGFGSYGSFNN 1080
DB 1013 APKYEKPTNVWNAITGGTSLNNGSNASLYGTSGAGVDAYLNGQVEAIVGGFGSYGSFNN 1072
QY 1081 RANSLNNGANNNTNFGYSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNQSXYHLA 1140
DB 1073 QANSLNNGANNNTNFGYSRIFANQHEFDEFAQAGLSDQSSLNFKSALLQDLNQSXYHLA 1132
QY 1141 YSAATRASGYDFAFPRNALVLKPSVGVNHLGSTNFKSNSTNOVALKNGSSSHLFNA 1200
DB 1133 YSATTRASGYDFAFPRNALVLKPSVGVNHLGSTNFKSNSTNOVALKNGSSSHLFNA 1192
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFYAHVGSNNAASLNTFKVNAARNPLNTHARVMMGE 1260
DB 1193 NANVEARYYYGDTSYFYMNAGVLQEFYAHVGSNNDVSLNTFKVNAARNPLNTHARVMMGE 1252
QY 1261 LKLAKEVFLNLGVVYVHLNLSNIGHFASNLGMRYSF 1296
DB 1253 LKLAKEVFLNLGVVYVHLNLSNIGHFASNLGMRYSF 1288

SEQUENCE FROM N.A.
RC STRAIN=F73;
RX MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori.";
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL: AF049652; AAD04289.1; -;
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca; 1.
SQ SEQUENCE 1296 AA; 139862 MW; 72469BEE8000BFDA CRC64;
```

Query Match 92.3%; Score 6154; DB 2; Length 1296;
Best Local Similarity 91.5%; Pred. No. 7.2e-228;
Matches 1186; Conservative 56; Mismatches 54; Indels 0; Gaps 0;

```
QY 1 MEIQOTHRKINRPLVSLALVGVLSITPQOASHAAFTTVIIPAIYVGGIATGAVGTVSG 60
DB 1 MEIQOTHRKINRPLVSLALVGVLSITPQOASHAAFTTVIIPAIYVGGIATGAVGTVSG 60
61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKKEYDLYBSLLSSKIDGGWDMGNAAR 120
61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPKKEYDLYBSLLSSKIDGGWDMGNAAR 120
```

Qy	121	HYWYKGGQONKLEVDMDKDAVGTYYTL	SLGRNFTGGDLVNMOKATLRLGQFN	NGNSFTSYKD	180
Db	121	HYWYKGGQONKLEVDMDKDAVGTYYTL	SLGRNFTGGDLVNMOKATLRLGQFN	NGNSFTSFKD	180
Qy	181	SADRTTRVDNNAKNI	SDNFVEINNRVSGAGRKASSTVLT	LOASEGITS	SDKNAEISLYD 240
Db	181	SADRTTRVDNNAKNI	SDNFVEINNRVSGAGRKASSTVLT	LOASEGITS	SDKNAEISLYD 240
Qy	241	GATLNLASSSVKLMGNVMMGRLOQV	GAYLAPSYSTINTSKVTGEVNFNHL	TGVDKNAQA	300
Db	241	GATLNLASSSVKLMGNVMMGRLOQV	GAYLAPSYSTINTSKVTGEVNFNHL	TGVDKNAQA	300
Qy	301	GIANKNTNIGTLDLW	SAGLNIIAPPEGYKDKPNT	PSQSGAKNDK	NSAKNDKQESS 360
Db	301	GIANKNTNIGTLDLW	SAGLNIIAPPEGYKDKPNT	PSQSGAKNDK	NSAKNDKQESS 360
Qy	361	QNNSTQVIPPNSAOKTE	VOPTQVIDGPFAGGKDTVVN	INRINTNADG	IRVGGFKASL 420
Db	361	QNNSTQVIPPNSAOKTE	VOPTQVIDGPFAGGKDTVVN	INRINTNADG	IRVGGFKASL 420
Qy	421	TTNAHLHIGKGGVNL	SNOASGRSLIVENLTGNTVDG	PLRVNNOVG	YALAGSSANFEF 480
Db	421	TTNAHLHIGKGGVNL	SNOASGRSLIVENLTGNTVDG	PLRVNNOVG	YALAGSSANFEF 480
Qy	481	KAGTDTKNGTATFN	NDISLGRFVNLKDAHTANPKG	IDTGN	GGFNTLDFSGVTDKVNINK 540
Db	481	KAGTDTKNGTATFN	NDISLGRFVNLKDAHTANPKG	IDTGN	GGFNTLDFSGVTDKVNINK 540
Qy	541	LITASTNVAKNFNI	NELIVKTNIGISGEYTHFSE	DIGSOSRINT	VRLETCRSLFSGGV 600
Db	541	LITASTNVAKNFNI	NELIVKTNIGISGEYTHFSE	DIGSOSRINT	VRLETCRSLFSGGV 600
Qy	601	KFKGGERLVIDEFY	SPWNTFDARNIKVEITNK	KLAFPGQSP	MGTSKLMFNLLTQNA 660
Db	601	KFKGGERLVIDEFY	SPWNTFDARNIKVEITNK	KLAFPGQSP	MGTSKLMFNLLTQNA 660
Qy	661	VMDYSQSNLTIGD	FTNNOGTINLVGRGKVATL	SVGNAAMFN	NDIDISATGYKPLI 720
Db	661	VMDYSQSNLTIGD	FTNNOGTINLVGRGKVATL	SVGNAAMFN	NDIDISATGYKPLI 720
Qy	721	KINSAQDLINKTH	VLKAKIIGYGVSTGTNG	ISNVLNLEO	PKERLALYNNNRMDTCV 780
Db	721	KINSAQDLINKTH	VLKAKIIGYGVSTGTNG	ISNVLNLEO	PKERLALYNNNRMDTCV 780
Qy	781	VRNTDDIKACGMA	IGDQSMVNNPDNYKYL	IGKAWKNIG	ISKTANGSKISVYLGNSPTPE 840
Db	781	VRNTDDIKACGMA	IGDQSMVNNPDNYKYL	IGKAWKNIG	ISKTANGSKISVYLGNSPTPE 840
Qy	841	NGGNTNLPNTN	SNARSANALAOAPPAP	QATPNLVAIN	QHDGFTIESVFELANRSK 900
Db	841	NGGNTNLPNTN	SNARSANALAOAPPAP	QATPNLVAIN	QHDGFTIESVFELANRSK 900
Qy	901	DIDTLYANSAQGR	DLLOTLIDSHDAGYARK	MDATANEIT	KOLATATTLNNAISLE 960
Db	901	DIDTLYANSAQGR	DLLOTLIDSHDAGYARK	MDATANEIT	KOLATATTLNNAISLE 960
Qy	961	HKTSGGLQTL	SLSNAMILNSRLVNSRR	HTNHDISFAKRL	QALQDKQFASLESAAEVLQF 1020
Db	961	HKTSGGLQTL	SLSNAMILNSRLVNSRR	HTNHDISFAKRL	QALQDKQFASLESAAEVLQF 1020
Qy	1021	APKYKPTN	WANAIGTSLNNGSN	ASLYGTSAGVD	AYLNGOVEAIVGGFGSYGYSFNN 1080
Db	1021	APKYKPTN	WANAIGTSLNNGSN	ASLYGTSAGVD	AYLNGOVEAIVGGFGSYGYSFNN 1080
Qy	1081	RANSLNGANTN	PGVYSRIFANOHED	FEAQALGDSQ	SLNFKSALLDOLNOSHYLA 1140
Db	1081	RANSLNGANTN	PGVYSRIFANOHED	FEAQALGDSQ	SLNFKSALLDOLNOSHYLA 1140
Qy	1141	YSAATRASGYD	FAFFRNALVLPKPSV	GVSYNHLGSTN	FKNSNTNQVALKNGSSQHLFNA 1200
Db	1141	YSAATRASGYD	FAFFRNALVLPKPSV	GVSYNHLGSTN	FKNSNTNQVALKNGSSQHLFNA 1200
Qy	1201	SANVEARYYYG	DTSYFYMNAGVLQEF	FAHVGSNNA	ASLNTFKVNARNPLNTHARVMMGE 1260

Db	1201	SANVEARYYYG	DTSYFYMNAGVLQEF	ANFGSSNAV	SLNTFKVNARNPLNTHARVMMGE 1260
Qy	1261	LKLAKEVFLNL	GVVYVYLNHLNLSNIG	HFASNLGMYSF	1296
Db	1261	LKLAKEVFLNL	GVVYVYLNHLNLSNIG	HFASNLGMYSF	1296
RESULT	9				
Q92H08					
ID	Q92H08	PRELIMINARY;	PRT;	1291	AA.
AC	Q92H08;				
DT	01-MAY-1999	(Tremblrel. 10, Created)			
DT	01-MAY-1999	(Tremblrel. 10, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	VACUOLATING	CYTOTOXIN.			
GN	VACA.				
OS	Helicobacter pylori	(Campylobacter pylori).			
OC	Bacteria; Proteobacteria;	epsilon subdivision; Helicobacter group;			
OC	Helicobacter.				
OX	NCBI_TaxID=210;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=F52;				
RC	MEDLINE=98453456;	PubMed=9780260;			
RA	Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y., Kuriyama M.;				
RA	"Full-length sequence analysis of the vacA gene from cytotoxic and noncytotoxic Helicobacter pylori."				
RT	J. Infect. Dis. 178:1391-1398(1998).				
RL	EMBL: AF049631; AAD04269.1; -				
DR	InterPro: IPR003842; Vaca.				
DR	Pfam: PF02691; Vaca; 1.				
DR	SEQUENCE	1291 AA; 139473 MW; 398037E	FEF290254	CRC64;	
Qy					
Qy	1	MELOOTHRKINRPLVSLALV	ALVSLITPQOSHAAFFTVIIP	ALVGGTATGTVGTSGL 60	
Db	1	MELOOTHRKINRPLVSLALV	ALVSLITPQOSHAAFFTVIIP	ALVGGTATGTVGTSGL 60	
Qy	61	LSWGLKQAEANKTPDK	PKVWRIQAGKGFNEFPNKEYDLY	RSLLSSKIDGWDGNAAR 120	
Db	61	LSWGLKQAEANKTPDK	PKVWRIQAGKGFNEFPNKEYDLY	RSLLSSKIDGWDGNAAR 120	
Qy	121	HYWYKGGQONKLEVD	MDKDAVGTYYTL	SLGRNFTGGDLVNMOKATLRLGQFN	NGNSFTSYKD 180
Db	121	HYWYKGGQONKLEVD	MDKDAVGTYYTL	SLGRNFTGGDLVNMOKATLRLGQFN	NGNSFTSFKD 180
Qy	181	SADRTTRVDNNAKNI	SDNFVEINNRVSGAGRKASSTVLT	LOASEGITS	SDKNAEISLYD 240
Db	181	SADRTTRVDNNAKNI	SDNFVEINNRVSGAGRKASSTVLT	LOASEGITS	SDKNAEISLYD 240
Qy	241	GATLNLASSSVKLMGNVMMGRLOQV	GAYLAPSYSTINTSKVTGEVNFNHL	TGVDKNAQA	300
Db	241	GATLNLASSSVKLMGNVMMGRLOQV	GAYLAPSYSTINTSKVTGEVNFNHL	TGVDKNAQA	300
Qy	301	GIANKNTNIGTLDLW	SAGLNIIAPPEGYKDKPNT	PSQSGAKNDK	NSAKNDKQESS 360
Db	301	GIANKNTNIGTLDLW	SAGLNIIAPPEGYKDKPNT	PSQSGAKNDK	NSAKNDKQESS 360
Qy	361	QNNSTQVIPPNSAOKTE	VOPTQVIDGPFAGGKDTVVN	INRINTNADG	IRVGGFKASL 420
Db	361	QNNSTQVIPPNSAOKTE	VOPTQVIDGPFAGGKDTVVN	INRINTNADG	IRVGGFKASL 420
Qy	421	TTNAHLHIGKGGVNL	SNOASGRSLIVENLTGNTVDG	PLRVNNOVG	YALAGSSANFEF 480
Db	421	TTNAHLHIGKGGVNL	SNOASGRSLIVENLTGNTVDG	PLRVNNOVG	YALAGSSANFEF 475
Qy	481	KAGTDTKNGTATFN	NDISLGRFVNLKDAHTANPKG	IDTGN	GGFNTLDFSGVTDKVNINK 540


```
Db 476 KAGVDTKNGTIAFNNNISLGRVNLKASHTVNFKDDIDTNGGNTLDFSGVTNKVNINK 535
Qy 541 LITASTNVAVKFNINELIVKNGTISVGEYTHFSEDIGSQSRINTVRLTGTGRSLFSGGV 600
Db 536 LITASTNVAIKFNINELIVKNGTISVGEYTHFSEDIGSQSRINTVRLTGTGRSLFSGGV 595
Qy 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPNA 660
Db 596 KFKGGEKLVINDFYAPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPNA 655
Qy 661 VMDYSQFSNLTTOGDFINNQTINYLVRGKGVATISVGNAAAMFNNDIDSATGYKPLI 720
Db 656 VMDYSQFSNVTIQGNFINNQGTINYLVRGNTIETLVGNAAVMFNNNDIDSATGYKPLI 715
Qy 721 KINSADLLKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 716 KINSADLLKNTHEVLLKAKIIGYENASLGTSISNANIEQFNERLALYNNNRMDTCV 775
Qy 781 VRNTDDIKACGMAIGDQSMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVIGNSTPTE 840
Db 776 VRNTDDIKACGMAIGDQAMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVIGNATPAE 835
Qy 841 NGCNTNTLPTNTTSNARSANNAQAQPSATPNLVAIINQHDFTGTTIESVFELANRSK 900
Db 836 NGCNTNTLPTNTTNARFASALIKNAPPAQTSATPNLVAIINKHDFTGTTIESVFELANRSK 895
Qy 901 DDTLYANGAGRODLQTLTLLSDHDAGYARKWIDATSAEITKOLNTATTTLNNTASLE 960
Db 896 DDTLYANGAGRODLQTLTLLSDHNGYARTWIDATSAEITKOLNTATDNLNTASLE 955
Qy 961 HKTSGLOTLSSLNAMILNRLNRRHTNHIDSFARKLQALKDQKFALESAAEVLQYF 1020
Db 956 HKTSGLOTLSSLNAMILNRLNRRHTNNIDSFARKLQALKDQKFALESAAEVLQYF 1015
Qy 1021 APKYEKPTNVWANAIGGTSLNNGSNALYGTSGAGVDAYLNGOVEALVGGFGSYGSSPNN 1080
Db 1016 APKYEKPTNVWANAIGGASLNNGSGGNTSLYGTSGAGVDAYLNEKVEALVGGFGSYGSSPNN 1075
Qy 1081 RANSLNGANNPTNFGYSRIFANQHEFDEAOCALGSDQSLLNFKSALLQDLNQSRYHLA 1140
Db 1076 QANSLNGANNANFGYSRIFANRHEFDEAOCALGSDQSLLNFKSALLQDLNQSRYHLA 1135
Qy 1141 YSAATRASGYDYDFARNALVKPSVGSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1136 YSAATRASGYDYDFARNALVKPSVGSYNHLGSTNFPESNTHKTALKANGASSQHLFNA 1195
Qy 1201 SANVEARYYGDTSFYFNAGVLQEFARHVGSNNAASLNTFFKYNAARNPLNTHARVMMGE 1260
Db 1196 SANVEARYYGDTSFYFNAGVLQEFARHVGSNNAASLNTFFKYNAARNPLNTHARVMMGE 1255
Qy 1261 LKLAKEVFLNLGVVYHLNLSINIGHFASNLGMRYSF 1296
Db 1256 LKLAKEVFLNLGVYHLNLSINIGHFASNLGMRYSF 1291
RESULT 10
Q9ZHT8
ID Q9ZHT8 PRELIMINARY; PRT: 1291 AA.
AC Q9ZHT8;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VACUOLATING CYTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F64;
RX MEDLINE=98453456; PubMed=9780260;
```

```
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori.";
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL; AF049647; AAD04282.1; -.
DR InterPro; IPR003842; VACA.
DR Pfam; PF03691; Vaca; 1.
SQ SEQUENCE 1291 AA; 139395 MW; 4B5785B5BEE7934 CRC64;
```

Query Match 91.9%; Score 6128.5; DB 2; Length 1291;
Best Local Similarity 91.2%; Pred. No. 6.8e-227;
Matches 1182; Conservative 55; Mismatches 54; Indels 5; Gaps 1;

```
Qy 1 MEIQOQTHRKINRPLVSLALVGVLSITPOOSHAAFTTTVIIIPAIVGGIATGATVAVTSGV 60
Db 1 MEIQOQTHRKINRPLVSLALVGVLSITPOKSHAAFTTTVIIIPAIVGGIATGATVAVTSGV 60
Qy 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYRSLSSKIDGGWDMGNAAR 120
Qy 121 HYWVKGQOQNKLEVDMKDAVGTYYLSGLRNTFGDLDVNMOKATLRLGQFNGNSFTSYKD 180
Db 121 HYWVKGQOQNKLEVDMKDAVGTYYLSGLRNTFGDLDVNMOKATLRLGQFNGNSFTSYKD 180
Qy 181 SADRTTRVDNFNAKNTSIDNFEVNNRVGSGAGRKASSTVLTLOASGEGITSKNAEISLYD 240
Db 181 SADRTTRVDNFNAKNTSIDNFEVNNRVGSGAGRKASSTVLTLOASGEGITSKNAEISLYD 240
Qy 241 GATNLASSSVKLMGNVMGRLOQYVAYLAPSYSTINTSKVTGEVNFNLTIVGDKNAAQA 300
Db 241 GATNLASSSVKLMGNVMGRLOQYVAYLAPSYSTINTSKVTGEVNFNLTIVGDKNAAQA 300
Qy 301 GIANKTKTIGTLDLWQSAAGNIAPPEGGYKDKPNTTPSQGAKNDKKNESAKNDKQES 360
Db 301 GIANKTKTIGTLDLWQSAAGNIAPPEGGYKDKPNTTPSQGAKNDKKNESAKNDKQES 358
Qy 361 QNNSQVTLNPNPNSAQKTEVQPTVIDGPFAGKDPVWVNRINTNADGTRVGGFKASL 420
Db 359 ---SNTQVTLNPNPNSAQKTEVQPTVIDGPFAGKDPVWVNRINTNADGTRVGGFKASL 415
Qy 421 TTNAHLHGKGVNLNSQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 480
Db 416 TTNAHLHGKGVNLNSQASGRSLIVENLTGNITVDGPLRVNNQVGGYALAGSSANFEF 475
Qy 481 KAGTDTKNGTATFNNDISLGRFVNLKVDATNFNGKIDGTGNGGFNTLDFSGVTDKYNINK 540
Db 476 KAGVDTKNGTATFNNNISLGRFVNLKASHTVNFKDDIDTNGGNTLDFSGVTDKYNINK 535
Qy 541 LITASTNVAVKFNINELIVKNGTISVGEYTHFSEDIGSQSRINTVRLTGTGRSLFSGGV 600
Db 536 LITASTNVAIKFNINELIVKNGTISVGEYTHFSEDIGSQSRINTVRLTGTGRSLFSGGV 595
Qy 601 KFKGGEKLVIDEFYSPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPNA 660
Db 596 KFKGGEKLVINDFYAPWNYFDARNIKNVEITNKLAFGPQGSPPWGTAKLMFNLLTGPNA 655
Qy 661 VMDYSQFSNLTTOGDFINNQTINYLVRGKGVATISVGNAAAMFNNDIDSATGYKPLI 720
Db 656 VMDYSQFSNVTIQGNFINNQGTINYLVRGNTIETLVGNAAVMFNNNDIDSATGYKPLI 715
Qy 721 KINSADLLKNTHEVLLKAKIIGYGNVSTGTNGISNVNLEEQFKERLALYNNNRMDTCV 780
Db 716 KINSADLLKNTHEVLLKAKIIGYENASLGTSISNANIEQFNERLALYNNNRMDTCV 775
Qy 781 VRNTDDIKACGMAIGDQSMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVIGNSTPTE 840
Db 776 VRNTDDIKACGMAIGDQAMVNNPDNYKYILIGKAWKNIGISKTAGSKISVYVIGNATPAE 835
Qy 841 NGCNTNTLPTNTTSNARSANNAQAQPSATPNLVAIINQHDFTGTTIESVFELANRSK 900
Db 836 NGCNTNTLPTNTTNARFASALIKNAPPAQTSATPNLVAIINKHDFTGTTIESVFELANRSK 895
```

```
Db 836 NGGNTTLPNTNATNNARFASVALIKNAPFAQTNPMLVAINKHDFGTIESVFELANRSK 895
Qy 901 DIDTLYANSAQGRDLLQTLIDSHDAGYARKMIDATSANETIKOLNTATTTLANNIASLE 960
Db 896 DIDTLYAHSGAQRDLLQTLIDSHDAGYARTMIDATSANETIKOLNTATDALNNIASLE 955
Qy 961 HKTSGLQTLSLSNAMILNSRLVNLSSRHTNHDSPAKRLQALKDQKFALESAAEVLVYQF 1020
Db 956 HKTSGLQTLSLSNAMILNSRLVNLSSRHTNHDSPAKRLQALKDQKFALESAAEVLVYQF 1015
Qy 1021 APKYEKPTNWANNAIGTSLNNGSNASLYGTSAGVDAYLNCQVEAIVGGFSGYSFNN 1080
Db 1016 APKYEKPTNWANNAIGTSLNNGSGNTSLYGTSAVDAYLNEKVEAIVGGFSGYSFNN 1075
Qy 1081 RANSLNGANTNFVYSRIFANQHEDFEAQAGALGDSQSLNFKSALLDLNQSYYLA 1140
Db 1076 QANSLNGANNANFVYSRIFANRHEDFEAQAGAVGDSQSLNFKSALLDLNQSYYLA 1135
1141 YSAATRASGYDFAFFRNALVLPKPSVGSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1136 YGAATRASGYDFAFFRNALVLPKPSVGSYNHLGSTNFKSNSTNKTALKNGASSQHLFNA 1195
Qy 1201 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Db 1196 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGGE 1255
Qy 1261 LKLAKEVFLNLGVYLLHNLISNIGHFASNLGMYSF 1296
Db 1256 LKLAKEVFLNLGVYLLHNLISNIGHFASNLGMYSF 1291
RESULT 11
Q9ZHT7 PRELIMINARY; PRT; 1296 AA.
AC Q9ZHT7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
RT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=F68;
MEDLINE=98453456; PubMed=9780260;
Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
noncytotoxic Helicobacter pylori.";
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL; AF049648; AAD04283.1; -.
DR InterPro; IPR003842; VACA.
DR Pfam; PF02631; Vaca; 1.
SQ SEQUENCE 1296 AA; 140049 MW; E66421537383F709 CRC64;
```

```
Query Match 91.88; Score 6122; DB 2; Length 1296;
Best Local Similarity 90.78; Pred. No. 1.2e-226;
Matches 1176; Conservative 65; Mismatches 55; Indels 0; Gaps 0;
Qy 1 MEIOTHRKINRPLVSLALVGLVSIIPQOSHAAFFTTVIIPALVGGIATGAVGTVSGL 60
Db 1 MEIOTHRKINRPLVSLALVGLVSIIPQOSHAAFFTTVIIPALVGGIATGAVGTVSGL 60
Qy 61 LSWGLKQAEAEANKTPDKPKVRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGNGAAR 120
Db 61 LSWGLKQAEAEANKTPDKPKVRIQAGKGFNEFPNKEYDLYRSLSSKIDGWDGNGAAR 120
Qy 121 HWYKGGQMKLEVDMDKDAVGTYKLSGLRNFTGGDLVDNMOKATLRLGQFNGNSFTSKD 180
Db 121 HWYKGGQMKLEVDMDKDAVGTYKLSGLRNFTGGDLVDNMOKATLRLGQFNGNSFTSKD 180
```

```
Db 121 HWYKGGQMKLEVDMDKDAVGTYKLSGLRNFTGGDLVDNMOKATLRLGQFNGNSFTSKD 180
Qy 181 SADRTTRVDFNAKNISIDNFEINNRVSGAGRKASSTVLTLOASEGITSBKNAEISLYD 240
Db 181 GTNRTTRVDFNAKNILLDNFEINNRVSGAGRKASSTVLTLOASEGITSBKNAEISLYD 240
Qy 241 GATLNLASSSYKLMGNVMMGRLOVYGAYLAPSYSTINTSKVTGEVNFNHLTVGDKNAQA 300
Db 241 GATLNLASNSYKLMGNVMMGRLOVYGAYLAPSYSTINTSKVTGEVNFNHLTVGDRNAQA 300
Qy 301 GIANKKTNIGTLDLWOSAGLNIITAPPEGGYKDKPNWTPSOSGAKNDKNSAKNDKQESS 360
Db 301 GIATSKTYITGTLWOSAGLNIITAPPEGGYKDKPNWTPSOSGAKNDKNSAKNDKQESS 360
Qy 361 QNNSNTQVINPPNSAQKTEVQPTQIDGPPAGGKDTVVYINRINTNADGTIRVGGFKASL 420
Db 361 QNNSNTQVINPPNSQKTEIQPTQIDGPPAGAKDTVWYINRINTNADGTIKVGGYTASL 420
Qy 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDPGLRVNNQVGGYALAGSAPFEF 480
Db 421 TTNAHLNIGKGGVNLNQASGRSLIVENLTGNTVEGALRVNNQVGGYALAGSAPFEF 480
Qy 481 KAGTDTKNGTATFNNDISLGRFVNLKVDATANFKGIDTGNCGGFNTLDFSGVTDKVINIK 540
Db 481 KAGVDTKNGTATFNNDISLGRFVNLKASHTVNEFKDIDTGNCGGFNTLDFSGVTNKVINIK 540
Qy 541 LITASTNVAVKFNINELIVKTNIGISVGEYTHFSEIDIGSQSRINTVRLTGTSLFSGGV 600
Db 541 LITASTNVAKFNINELIVKTNIGISVGEYTNFSEIDIGNQSRINTVRLTGTTRIYSGGV 600
Qy 601 KFKGGEKLVIDEFYSPWNYFDARNIKVEITNKLAFEPQSGSPWGTSKLPMNLTLCQNA 660
Db 601 KFKGGEKLVINDEFYAPWNYFDARNIKVEITNKLAFEPQSGSPWGTAKLPMNLTLCGNA 660
Qy 661 VMDYSQFSLTIQGFNNQGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFYKPLI 720
Db 661 VMDYSQFSLTIQGFNNQGTINYLVRGGNIETLSVGNAAAMFNNDIDSATGFYKPLI 720
Qy 721 KINSAQDLIKNTEHVLKAKIIGYGNVSTGTNGISNVNLEOFPKRLALYNNNRMDTCV 780
Db 721 KINSAQDLIKNKEHVLKAKIIGYGNVSLGTNSISNANLEOFPKRLALYNNNRMDTCV 780
Qy 781 VRNTDDIKACGMAITGDSQSVNPNPNYKYLIGKAWKNIGISKANGSKISVYLLGNSTPTE 840
Db 781 VRNTDDIKACGMAITGDSQSVNPNPNYKYLIGKAWKNIGISKANGSKISVYLLGNATPPE 840
Qy 841 NGGNTTLPNTTNSARSNANALQAQNPAPQATPNLVAINHDFGTIESVFELANRSK 900
Db 841 NGGNTTLPNTTNNARFASVALIKNAPFAQTNPMLVAINKHDFGTIESVFELANRSK 900
Qy 901 DIDTLYANSAQGRDLLQTLIDSHDAGYARKMIDATSANETIKOLNTATTTLANNIASLE 960
Db 901 DIDTLYAHSGAQRDLLQTLIDSHNAGYARTMIDATSANETIKOLNANSALNNIASLE 960
Qy 961 HKTSGLQTLSLSNAMILNSRLVNLSSRHTNHDSPAKRLQALKDQKFALESAAEVLVYQF 1020
Db 961 HKTNGLQTLSLSNAMILNSRLVNLSSRHTNHDSPAKRLQALKDQKFALESAAEVLVYQF 1020
Qy 1021 APKYEKPTNWANNAIGTSLNNGSNASLYGTSAGVDAYLNCQVEAIVGGFSGYSFNN 1080
Db 1021 APKYEKPTNWANNAIGTSLNNGSGNTSLYGTSAVDAYLNEKVEAIVGGFSGYSFNN 1080
Qy 1081 RANSLNGANTNFVYSRIFANQHEDFEAQAGALGDSQSLNFKSALLDLNQSYYLA 1140
Db 1081 QANSLNGANNANFVYSRIFANRHEDFEAQAGAVGDSQSLNFKSALLDLNQSYYLA 1140
Qy 1141 YSAATRASGYDFAFFRNALVLPKPSVGSYNHLGSTNFKSNSTNOVALKNGSSQHLFNA 1200
Db 1141 YGAATRASGYDFAFFRNALVLPKPSVGSYNHLGSTNFKSNSTNKTALKNGASSQHLFNA 1200
Qy 1201 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
Db 1201 SANVEARYYYGDTSYFYMNAGVLQEFARHVGSSNNAASLNTFKVNAARNPLNTHARVMMGGE 1260
```



```

Qy 1261 LKLAKEVFLNLGVVYVYHNLNISNIGHFASNLGMRYSF 1296
Db 1261 LKLAKEVFLNLGVYIYHNLNISNIGHFASNLGMRYSF 1296

RESULT 12
O34108 PRELIMINARY; PRT; 1296 AA.
ID 034108
AC 034108;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
OS VACA.
GN Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
ON NCBI_Taxid=210;
RX [1]
RP SEQUENCE OF 517-760 FROM N.A.
RA STRAIN=F71;
RA MEDLINE=97339580; PubMed=9196179;
RA Ito Y., Azuma T., Ito S., Miyaji H., Hirai M., Yamazaki Y., Sato F.,
RA Kato T., Kohli Y., Kuriyama M.;
RA "Analysis and typing of the vacA gene from caga-positive strains of
RT Helicobacter pylori isolated in Japan.,"
RT J. Clin. Microbiol. 35:1710-1714(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=F71;
RA MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RA "Full-length sequence analysis of the vacA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori.,"
RT J. Infect. Dis. 178:1391-1398(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=F71;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF071096; AAC77451.1; -.
SQ SEQUENCE 1296 AA; 139786 MW; F45E2449C811C2BB CRC64;

Query Match 91.8%; Score 6122; DB 2; Length 1296;
Best Local Similarity 90.8%; Pred. No. 1.2e-226;
Matches 1177; Conservative 61; Mismatches 58; Indels 0; Gaps

1 MELOOQTHRKINRPLVSLALVGLSVITPOQSHAAFTTIIPIAVGCIATGAVGTVSGL 60
11 MELOOQTHRKINRPLVSLALVGLSVITPOQSHAAFTTIIPIAVGCIATGAVGTVSGL 60
11 MELOOQTHRKINRPLVSLALVGLSVITPOQSHAAFTTIIPIAVGCIATGAVGTVSGL 60

61 LSWGLQKQAEANKTPDKPDKVWRIQAGKFNPEPNKEYDLYRSLSSKIDGGWDGNAAR 120
61 LSWGLQKQAEANKTPDKPDKVWRIQAGKFNPEPNKEYDLYRSLSSKIDGGWDGNAAR 120
61 LGWGLQKQAEANKTPDKPDKVWRIQAGKFNPEPNKEYDLYRSLSSKIDGGWDGNAAR 120

121 HYWVGQGNKLEVDMMKDAVGTVTLSGLRNFTGGDLVNNMQKATLRLGQFNNGNSTSYKD 180
121 HYWVGQGNKLEVDMMKDAVGTVKLSGLRNFTGGDLVNNMQKATLRLGQFNNGNSTSYKD 180
121 HYWVGQGNKLEVDMMKDAVGTVKLSGLRNFTGGDLVNNMQKATLRLGQFNNGNSTSYKD 180

181 SADRTTRVDNFNAKNISIDNFVEINNRVGSAGRKASSTVLTLOASGKITSDRNAEISLYD 240
181 SADRTTRVDNFNAKNISIDNFVEINNRVGSAGRKASSTVLTLOASGKITSDRNAEISLYD 240
181 SADRTTRVDNFNAKNISIDNFVEINNRVGSAGRKASSTVLTLOASGKITSDRNAEISLYD 240

241 GATNLINASSSVKLMGNVWGRQYVCAYLAPSVSTINTSKVTEVNFNHLTVGDKNAQA 300
241 GATNLINASSSVKLMGNVWGRQYVCAYLAPSVSTINTSKVTEVNFNHLTVGDKNAQA 300
241 GATNLINASSSVKLMGNVWGRQYVCAYLAPSVSTINTSKVTEVNFNHLTVGDKNAQA 300

301 GIANKKNTGTLDLWQAGLNIAPPEGYKDKPNPTPSQSAKNDKRNKSAKNDKQESS 360
301 GIANKKNTGTLDLWQAGLNIAPPEGYKDKPNPTPSQSAKNDKRNKSAKNDKQESS 360
301 GIANKKNTGTLDLWQAGLNIAPPEGYKDKPNPTPSQSAKNDKRNKSAKNDKQESS 360

```

01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN (FRAGMENT).
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHN5060D;
RA Ji X., Telford J.L., Burroni D., Guidotti S., Pagliaccia C.,
RA Rayrat J.M., Xu G., Rappuoli R.;
RT "Allelic variation of vacA gene in the Chinese Helicobacter pylori."
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF050328; AAF26510.1; -
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca; 1.
DR NON_TER 1291 1291
SEQUENCE 1291 AA; 139443 MW; 8C525ED343226392 CRC64;

Query Match 91.7%; Score 6120.5; DB 2; Length 1291;
Best Local Similarity 91.1%; Pred. No. 1.4e-226;
Matches 1181; Conservative 56; Mismatches 54; Indels 5; Gaps 1;

QY 1 MEIQOHRKINRPLVSLALVGLVSIITPQSHAAFFTTVIIPAIIVGGIATGTAAGTYSGL 60
DB 1 MELOOHRKINRPLVSLALVGLVSIITPQSHAAFFTTVIIPAIIVGGIATGTAAGTYSGL 60
QY 61 LSWGLKQAEANKTPDKPDKVWRIOAQKGFNEFPNKEYDLYRSLLSSKIDGGDWGNAAR 120
DB 61 LGWLKQAEANKTPDKPDKVWRIOAQKGFNEFPNKEYDLYRSLLSSKIDGGDWGNAAR 120
QY 121 HYWKGQONKLEVDKMDVAGTYTSLGRNFTGGDLVDNMOKATLRLQFNGNSFTSYKD 180
DB 121 HYWKGQONKLEVDKMDVAGTYTSLGRNFTGGDLVDNMOKATLRLQFNGNSFTSYKD 180
QY 181 SADRTTRVDFNAKISIDNFEINNRVSGAGRKASSTVLTQLASEGITSKNAEISLYD 240
DB 181 SADRTTRVDFNAKISIDNFEINNRVSGAGRKASSTVLTQLASEGITSKNAEISLYD 240
QY 241 GATNLASSSVKLMGNVWMLQVGLAYLAPSYTINTSKVTGVNFNHLTVGDKNAQA 300
DB 241 GATNLASSSVKLMGNVWMLQVGLAYLAPSYTINTSKVTGVNFNHLTVGDKNAQA 300
QY 301 GIANKTNTIGTLDWOSAGLNIIAPPEGGYKDKPNTTPSOGAKNDKNSAKNDKQESS 360
DB 301 GIASKKTYIGTLDWOSAGLNIIAPPEGGYKDKPNTTPSOGAKNDKNSAKNDKQOD-- 358
QY 361 QNNSNTQVPPNSAQKTEVQPTQVIDGPPAGGKDTVVNINRINTNADGTTIRVGGFKASL 420
DB 359 ---SNTQVPPNSQKTEIQPTQVIDGPPAGGKDTVVNINRINTNADGTTIRVGGFKASL 415
QY 421 TTNAAHLHGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNOVGYSALAGSANPEF 480
DB 416 TTNAAHLHGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNOVGYSALAGSANPEF 475
QY 481 KAGDTKNGTATFNNDISLGRFVNLKDAHTANPKGIDTNGGNTLDFSGVTDKVNINK 540
DB 476 KAGVDTKNGTATFNNDISLGRFVNLKDAHTANPKGIDTNGGNTLDFSGVTDKVNINK 535
QY 541 LITASTVAVKFNINELIVKNTGISVGEYTHFSDIGSQSRINTVLETRGTRSLFSGGV 600
DB 536 LITASTVAVKFNINELIVKNTGISVGEYTHFSDIGSQSRINTVLETRGTRSLFSGGV 595
QY 601 KFKGSEKLVIDFYYSPWNYFDARNIKNVEITNKLAFSGQSPWGTSKLMFNNTLQONA 660
DB 596 KFKGSEKLVIDFYYSPWNYFDARNIKNVEITNKLAFSGQSPWGTSKLMFNNTLQONA 655
QY 661 VMDYSQFSNLTIOGDFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFKPLI 720
DB 656 VMDYSQFSNLTIOGDFINNOGTINYLVRGGKVATLSVGNAAAMFNNDIDSATGFKPLI 715

QY 721 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEBQFKERLALYNNNRMDTCV 780
DB 716 KINSAQDLIKNTEHVLLKAKIIGYGNVSTGTNGISNVNLEBQFKERLALYNNNRMDTCV 775
QY 781 VRNTDDIKACMGATGDSQSVNPNPNYKYLIGKAWNKIGISKANGSKTSVYVLCNSTPTE 840
DB 776 VRNTDDIKACMGATGDSQSVNPNPNYKYLIGKAWNKIGISKANGSKTSVYVLCNATPAE 835
QY 841 NGGNTTNLPTNTTNSARSANNAQAAPPQSPATPNLVAIHOHDFGTIESVFELANRSK 900
DB 836 NGGNTTNLPTNTTNSARSANNAQAAPPQSPATPNLVAIHOHDFGTIESVFELANRSK 895
QY 901 DIDTLYANSQAQRDLLOTLIDSHDAGYAKKMTDATSANEITKQLTATTTLNIIASLE 960
DB 896 DIDTLYANSQAQRDLLOTLIDSHDAGYAKKMTDATSANEITKQLTATTTLNIIASLE 955
QY 961 HKTSGLQTLISLNAIILNSRLVLSRRHTNHDIFARLQALDKQKFALESAAEVLYQF 1020
DB 956 HKTSGLQTLISLNAIILNSRLVLSRRHTNHDIFARLQALDKQKFALESAAEVLYQF 1015
QY 1021 APKYKPTNWNANAIGGTSLNNGSNASLYGTSGVDAYLNGQVEAIVGGFSGYSGFNN 1080
DB 1016 APKYKPTNWNANAIGGTSLNNGSNASLYGTSGVDAYLNGQVEAIVGGFSGYSGFNN 1075
QY 1081 RANSLNGANNTPGVYSRIFANOHEDFEAQAQALGSDQSLNFKSALLQDLNOSYHYLA 1140
DB 1076 QANSLNGANNTPGVYSRIFANOHEDFEAQAQALGSDQSLNFKSALLQDLNOSYHYLA 1135
QY 1141 YSAATRASVYGDPAFFRNALVLPKPSGVSYNHLGSTNFKNSTNOVALKNGSSSOHLFNA 1200
DB 1136 YGAATRASVYGDPAFFRNALVLPKPSGVSYNHLGSTNFKNSTNOVALKNGSSSOHLFNA 1195
QY 1201 SANVEARYYYGDTSYFYMNAGVQLQEFARHVSNNAAALSTFKVNAARNPLNTHARVMGGE 1260
DB 1196 SANVEARYYYGDTSYFYMNAGVQLQEFARHVSNNAAALSTFKVNAARNPLNTHARVMGGE 1255
QY 1261 LKLAKEVFLNGVYVLLNLSNIGHFASNLGMRYSF 1296
DB 1256 LQLAKEVFLNGVYVLLNLSNIGHFASNLGMRYSF 1291

RESULT 14
Q9ZHT5 PRELIMINARY; PRT; 1291 AA.
AC Q9ZHT5;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VACUOLATING CYTOTOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC69;
RX MEDLINE=98453456; PubMed=9780260;
RA Ito Y., Azuma T., Ito S., Suto H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.;
RT "Full-length sequence analysis of the vacA gene from cytotoxic and
RL noncytotoxic Helicobacter pylori."
RL J. Infect. Dis. 178:1391-1398(1998).
DR EMBL: AF049649; AAD04285.1; -
DR InterPro: IPR003842; Vaca.
DR Pfam: PF02691; Vaca; 1.
DR SEQUENCE 1291 AA; 139363 MW; C568C500A2EBE8CF CRC64;

Query Match 91.7%; Score 6117.5; DB 2; Length 1291;
Best Local Similarity 91.2%; Pred. No. 1.8e-226;
Matches 1182; Conservative 56; Mismatches 53; Indels 5; Gaps 1;

QY 1 MEIQOTHRKINRPLVSLVGVALVITPQOQSHAAFFTTVIIPIAIVGGIATGTAAGTAVTSG 60
Db 1 MEIQOTHRKINRPLVSLVGVALLVITPQKSHAAFFTTVIIPIAIVGGIATGTAAGTAVTSG 60
QY 61 LSWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYBSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYBSLSSKIDGGWDMGNAAR 120
QY 121 HWYKGGQONKLEVDKMDAVGTYTTLGSLRNFPGDLDVNMOKATLRLGQFNGNSFTSYKD 180
Db 121 HWYKGGQONKLEVDKMDAVGTYTTLGSLRNFPGDLDVNMOKATLRLGQFNGNSFTSYKD 180
QY 181 SADRTTRVDFNAKNTSIDNFVEINNRVSGAGKASSTVLTILQASEGITSDKNABEISLYD 240
Db 181 SADRTTRVDFNAKNTSIDNFLEINNRVSGAGKASSTVLTILQASEGITSDKNABEISLYD 240
QY 241 GATNLASSSVKLMGVNMGRLOYVGYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
Db 241 GATNLASSSVKLMGVNMGRLOYVGYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
QY 301 GIANKKTNICTLDLQWQAGNIAPPEGGYKDKPNTPSQSGAKNDKNSAKNDKQESS 360
Db 301 GIASKKTYIGTLDLQWQAGNIAPPEGGYKDKPNTPSQSGAKNDKNSAKNDKQD-- 358
QY 361 QNNSTQVINPPNSAQKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Db 359 ---SNTQVINPPNSAQKTEIQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 415
QY 421 TTNAHLHIGKGVNLSNOAGSRSLIVENLTCNITVDGPLRVNNOVGGYALAGSSANFEF 480
Db 416 TTNAANLHIGKGVNLSNOAGSRSLVNLTCNITVEGALMVNNOVGGYALAGSSANFEF 475
QY 481 KAGDTKNGTATFNNDISLGRFVNLKVAHTANFKGIDTNGGFFNTLDFSGVTDKVNINK 540
Db 476 KAGVDTKNGTATFNNDISLGRFVNLKVAHTANFKGIDTNGGFFNTLDFSGVTDKVNINK 535
QY 541 LITASTNAVKNFNEILVKTNGTSVEGYTHFSBEDISQSRINTVRLTGTRSLRFSGV 600
Db 536 LITASTNAVKNFNEILVKTNGTSVEGYTHFSBEDISQSRINTVRLTGTRSLRFSGV 595
QY 601 KFKGGEKLVDFEYFSPNYPDARNIKVETNKNLAFQPGSPWGTSKLMFNNLTGQNA 660
Db 596 KFKGGEKLVDFEYFAPNYPDARNIKVETNKNLAFQPGSPWGTAKLMFNNLTGQNA 655
QY 661 VMDYSQFSLTIQGFINNQGTINLVRRGKVATLSVGNAAAMFNNDIDSATGYKPLI 720
Db 656 VMDYSQFSLTVIQGFINNQGTINLVRRGKVATLSVGNAAAMFNNDIDSATGYKPLI 715
QY 721 KINSAQDLINKNHEVLLKAKIIGYGNVSTGTNGISNVNLEQFKERLALYNNNNRMDTCV 780
Db 716 KINSAQDLINKNHEVLLKAKIIGYENASLGTNSISNANLIEQFNERLALYNNNNRMDTCV 775
QY 781 VRNTDDIKACGAIQDQWVNNPDNPKYILCKAWKNIGISKTAGSKTSVYVYLGHNSTPTE 840
Db 776 VRNTDDIKACGAIQDQWVNNPDNPKYILCKAWKNIGISKTAGSKTSVRYLGHNATFAE 835
QY 841 NGGNTNLPTNTTNSARNANLAQNAFPAQSPATPNLVAIINQHDFTGIESVFELANRSK 900
Db 836 NGGNTNLPTNATNARFASYALIKNAPPAQTNATPNLVAINKHDFGTIESVFELANRSK 895
QY 901 DIDTLVANSAGQGRDQLLTLLIDSHDAGYARKMIDATSAEITKQNTATTTLLNNIASLE 960
Db 896 DIDTLVANSAGQGRDQLLTLLIDSHDAGYARTMIDATSAEITKQNTATDALNNIASLE 955
QY 961 HKTSGLOTLSLNAMILNLSRLVNLRRHTNHDIDSAKRLQALKDQKFALESAAEVLQF 1020
Db 956 HKTSGLOTLSLNAMILNLSRLVNLRRHTNHDIDSAKRLQALKDQKFALESAAEVLQF 1015
QY 1021 APKYEKPTNVANATIGTSLNAGSNALYGTGSAGVDAYLNGQVEAIVGGFSGYGSFNN 1080
Db 1016 APKYEKPTNVANATIGTSLNAGSNALYGTGSAGVDAYLNGQVEAIVGGFSGYGSFNN 1075
QY 1081 RANSLNSGANTNFGVYSRIFANQHEFDEAQAALGSDQSSLNFKSALLQDLNQSYHYIA 1140

Db 1076 QANSTNSGANNANFGVYSRIFANRHEFDEAQAAGVSDQSSLNFKSALLQDLNQSYNYLA 1135
QY 1141 YSAATRASGYDXDFAFERNALVLKPSVGVSYNHLGSTNFKSNSTNOVALKNGSSOHLFNA 1200
Db 1136 YGAATRASGYDXDFAFERNALVLKPSVGVSYNHLGSTNFKSNSTNKTALKNGSSOHLFNA 1195
QY 1201 SANVEARYYYGDTSYFYMNAGVLQEFQFAHVGSSNAASLNTFKVNAARNPLNTHARVNMGE 1260
Db 1196 SANVEARYYYGDTSYFYMNAGVLQEFQFAHVGSSNAASLNTFKVNAARNPLNTHARVNMGE 1255
QY 1261 LKLAKEVFLNLGVVYVHLNLSINISGHIFASNLGMYSF 1296
Db 1256 LKLAKEVFLNLGVVYVHLNLSINISGHIFASNLGMYSF 1291
RESULT 15
O06021 PRELIMINARY; PRT; 1290 AA.
ID O06021;
AC O06021;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VACUOLATING CYTOFOXIN.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43526;
RA Ogura K., Maeda S., Kanai F.;
RT "Vacuolating cytotoxin".
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001338; AAB53868.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
SQ SEQUENCE 1290 AA; 139182 MW; 625406E581580578 CRC64;

Query Match 91.6%; Score 6113; DB 2; Length 1290;
Best Local Similarity 91.7%; Pred. No. 2.7e-226;
Matches 1188; Conservative 43; Mismatches 59; Indels 6; Gaps 2;
QY 1 MEIQOTHRKINRPLVSLVGVALVITPQOQSHAAFFTTVIIPIAIVGGIATGTAAGTAVTSG 60
Db 1 MEIQOTHRKINRPLVSLVGVALVITPQOQSHAAFFTTVIIPIAIVGGIATGTAAGTAVTSG 60
QY 61 LSWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYBSLSSKIDGGWDMGNAAR 120
Db 61 LSWGLKQAEAEANKTPDKPKVWRIQAGKGFNEFPNKEYDLYBSLSSKIDGGWDMGNAAR 120
QY 121 HWYKGGQONKLEVDKMDAVGTYTTLGSLRNFPGDLDVNMOKATLRLGQFNGNSFTSYKD 180
Db 121 HWYKGGQONKLEVDKMDAVGTYTTLGSLRNFPGDLDVNMOKATLRLGQFNGNSFTSYKD 180
QY 181 SADRTTRVDFNAKNTSIDNFVEINNRVSGAGKASSTVLTILQASEGITSDKNABEISLYD 240
Db 181 SADRTTRVDFNAKNTSIDNFLEINNRVSGAGKASSTVLTILQASEGITSDKNABEISLYD 240
QY 241 GATNLASSSVKLMGVNMGRLOYVGYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
Db 241 GATNLASSSVKLMGVNMGRLOYVGYLAPSYSTINTSKVTGEVNFHNLTVGDKNAQA 300
QY 301 GIANKKTNICTLDLQWQAGNIAPPEGGYKDKPNTPSQSGAKNDKNSAKNDKQESS 360
Db 301 GIASKKTYIGTLDLQWQAGNIAPPEGGYKDKPNTPSQSGAKNDKNSAKNDKQD-- 355
QY 361 QNNSTQVINPPNSAQKTEVQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 420
Db 356 QNNSTQVINPPNSAQKTEIQPTVIDGPFAGGKDTVVNINRINTNADGTIRVGGFKASL 415

QY 421 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGYALAGSSANPEF 480
Db 416 TTNAHLHIGKGGVNLNQASGRSLIVENLTGNTVDGPLRVNNQVGYALAGSSANPEF 475
QY 481 KAGDTKNGTATFNDISLGRFVNLKVDADHTANFKGIDTNGGPFNTLDFSCVTDKVNINK 540
Db 476 KAGDTKNGTATFNDISLGRFVNLKVDADHTANFKGIDTNGGPFNTLDFSCVTDKVNINK 535
QY 541 LITASTNVAVKFNINELIYKVTNGISVGEYTHFSEDIGSQSRINTVRLETGTRSLFSGGV 600
Db 536 LITASTNVAVKFNINELIYKVTNGISVGEYTHFSEDIGSQSRINTVRLETGTRSLFSGGV 595
QY 601 KFKGGEKLVIDEFYSPWNTFDARNIKNVEITNKLAFGPOGSPWGTSKLMFNNTLQONA 660
Db 596 KFKGGEKLVIDEFYSPWNTFDARNIKNVEITNKLAFGPOGSPWGTSKLMFNNTLQONA 655
QY 661 VMDYSQFSNLTIOGDFINNOGTINYLVRGKQVATLSVGNAAAMFNNDIDSATGCFYKPLI 720
Db 656 VMDYSQFSNLTIOGDFINNOGTINYLVRGKQVATLSVGNAAAMFNNDIDSATGCFYKPLI 715
QY 721 KINSAQDLIKNTEHVLKAKIIGYGVNVTGTNGISVNLBEOQFKERLALYNNNNRMDTCV 780
Db 716 KINSAQDLIKNTEHVLKAKIIGYGVNVTGTNGISVNLBEOQFKERLALYNNNNRMDTCV 775
QY 781 VRNTDDIKACOMAIGDOSMNNPDNYKYLICKAWKNIGISKGTANGSKISVYILGNSTPTE 840
Db 776 VRNTDDIKACOMAIGDOSMNNPDNYKYLICKAWKNIGISKGTANGSKISVYILGNSTPTE 835
QY 841 NGGNTNLTPTNTNSANNAQAAPFAOPSATPNLVAINOHDGCTIESVFELANRSK 900
Db 836 NGGNTNLTPTNTNSANNAQAAPFAOPSATPNLVAINOHDGCTIESVFELANRSK 894
QY 901 DIDPLYSANSAQGRDLLQTLIDSHDAGYVRQMDINTSTGEITKQLNAATTLNNIASLE 960
Db 895 DIDPLYSANSAQGRDLLQTLIDSHDAGYVRQMDINTSTGEITKQLNAATTLNNIASLE 954
QY 961 HKTSGLOTLSLNSAMILNSRLNLSRHTNIDSFARLQALKDQKFALESAAEVLYQF 1020
Db 955 HKTSGLOTLSLNSAMILNSRLNLSRHTNIDSFARLQALKDQKFALESAAEVLYQF 1014
QY 1021 APKYEKPTNVWANAIGTSLNNGSNASLYGTSGVDAYLNGOVEAIVGGFGSYGYSFNN 1080
Db 1015 APKYEKPTNVWANAIGTSLNNGSNASLYGTSGVDAYLNGOVEAIVGGFGSYGYSFNN 1074
QY 1081 RANSLNGANTNPGVYSRIFANQHEFDFAQAGALGSDQSSLNPKSALLQDLNQSYHYLA 1140
Db 1075 RANSLNGANTNPGVYSRIFANQHEFDFAQAGALGSDQSSLNPKSALLQDLNQSYHYLA 1134
QY 1141 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1200
Db 1135 YSAATRASGYDFAFFRNALVLPKPSGVSYNHLGSTNFKSNSTNQVALKNGSSSQHLFNA 1194
QY 1201 SANVEARYYGDTSYFYMNAGVQLQEFAGHAGVSNNAASLNTFKVNAARNPLNTHARVMGMGE 1260
Db 1195 SANVEARYYGDTSYFYMNAGVQLQEFAGHAGVSNNAASLNTFKVNAARNPLNTHARVMGMGE 1254
QY 1261 LKLAKEVFLNLGVVYHLNLSNIGHFASNLGMRYSF 1296
Db 1255 LKLAKEVFLNLGVVYHLNLSNIGHFASNLGMRYSF 1290